Wireless access network optimization for 5G

Gang, Jinwei

Awarding institution:
King's College London

Wireless Access Network Optimization for 5G

Jinwei Gang

Department of Informatics

King's College London

This dissertation is submitted for the degree of

Doctor of Philosophy

May 2019

Acknowledgements

I would like to express my sincere gratitude to Dr Vasilis Friderikos for accepting me to the PhD programme and for his continuous support, patience, motivation and inspiration. His guidance helped me in all the time of research. In addition, I would like to thank my family, who has always had my back and believed in me. I owe all the achievements to them. Lastly, I like to appreciate the help from all my friends in King's College London, who are great companions for me in London.

Abstract

The goal of expected 5th Generation (5G) wireless networks is to bring ultra high data rates, network throughput and service quality to mobile users in near future. Regarding to these expectations, the present virtualized and softwarized network architecture such as Cloud-Radio Access Network (C-RAN) and Software Defined Networks (SDN) is responsible for performing multiple key tasks such as allocating radio resources efficiently, providing Network Function Virtualization (NFV) and intelligent inter-network control/coordination, and flexibly sharing network resources among different network tenants and etc. Although these mentioned techniques and approaches already achieved certain progresses but also face various challenges because of the realistic engineering constraints and complicated network scenarios.

To this end, this thesis contributes a series of 5G wireless networks optimization frameworks and efficient algorithms tackling different network problems in both theoretical and practical ways. The specific works of the thesis include the following proposals: a deep research of control plane in future SDN architecture, which is capable to provide intelligent control functions. Furthermore, based on such SDN based architecture, a control plane signal optimization framework is defined and solved, which optimally reduces the potential handover signals and balances control load between multiple control planes in the SDN enabled networks.

On the other hand, by introducing network virtualization technologies, this thesis further proposes an optimization framework to flexibly share radio resources between network operators by using both intra-tenant and inter-tenant management entity. In addition, downlink signal transmission power control is also introduced to the framework to engage further resource reuse among tenants and achieve certain degrees improvement in energy efficiency during network operations. Due to the high complexity of the proposed optimization problem, various algorithms are proposed to solve this inter-tenant resource sharing problem efficiently, and numerical results demonstrate a satisfying improvements in various network performance.

Finally, a service rate aware resource sharing optimization framework is proposed to bring the previous inter-tenant sharing framework to a more realistic network architecturing level by considering fronthaul limitation and potential multi-cell coordinated resource transmission. Since the high complexity of this optimization problem, novel algorithm is also designed and numerical result demonstrates satisfying gains in a very complicated 5G networks environment.

Table of contents

List of figures

List of tables

List of Acronyms

BS: Base Station

BBU: Baseband Unit

C-RAN: Cloud Radio Access Network

CDF: Cumulative Density Function

C/U Split Network: Control Plane and User Plane Split Network

CoMP: Coordinated Multi-Point Communication

CPRI: Common Public Radio Interface

EE: Energy Efficiency

HetNet: Heterogeneous network

IRRA: Iterative Randomization Resource Allocation

InP: Infrastructure Provider

LX: Loose Coupling Inter-Tenant Resource Sharing

MME: Mobility Management Entity

MBS: Macro Base Station

MNO: Mobile Network Operator

MILP: Mixed Integer Linear Programming

MINLP: Mixed Integer Non-Linear Programming

NVS: Network Virtual Substrate

NVS-RB: Network Virtual Substrate-Resource Based

NFV: Network Function Virtualization

OFDMA: Orthogonal Frequency-Division Multiple Access

PPP: Poisson Point Process

PRB: Physical Resource Block

RRC: Radio Resource Control

RRH: Remote Radio Head

RPJA-h: Greedy based Resource and Power Joint Allocation

RPJA-adv: Advanced Resource and Power Joint Allocation

RA-h: Greedy based Resource Allocation

RA-adv: Advanced Resource Allocation

SBS: Small Base Station

SINR: Signal to Noise Ratio

SDN: Software Defined Network

SR: Static Reservation

SP: Service Provider

SDM-C: Software Defined Mobile Common Controller

SDM-X: Software Defined Mobile Inter-Tenant Controller

TX: Tight Coupling Inter-Tenant Resource Sharing

UE: User Equipment

3GPP: 3rd Generation Partnership Project

5G: 5th Generation Wireless Network

4G/LTE: 4th Generation Wireless Network

3GPP: 3rd Generation Partnership Project

Chapter 1

INTRODUCTION

1.1 Motivation

In order to plan, develop and deploy future 5G wireless networks, there is a plethora of technologies, architectural elements and algorithms that will be required to propel us from the present network architecture. Among those, prominent position have the concepts of cell densification and multi-tier network architecture which can be deemed as the two main stream technologies to allow for ultra high achievable data rates and various service types. It is envisioned that they will allow us to move towards high cell density and efficient heterogeneity in wireless access and improve overall service quality and efficiency. An overarching future 5G architectural paradigm shift is taking place towards the physical and/or logical split of the Control and User planes (C/U planes), which can be deemed as concept coming from the nominal SDN based architecture. However, since user plane performance is inevitably intertwined with that of the control plane, special attention should be paid to cases of congestion episodes, where limited control plane capacity might adversely affect overall network performance. To alleviate such potential control signal congestion, the thesis firstly provides a novel functional decomposed architecture and a network-wide optimization framework for logical C/U split network with emphasis on control plane load balancing and C-plane load reduction by taking into account inter macro controller handover and back-hauling limitations.

Meanwhile, to allow network sharing among different operators, network virtualization was introduced to the core of 5G atchitecturing and well researched by academia and industry for the last decade. With this capacity, multiple network operators are able to share a common network infrastructure, known as multi-tenancy/ operator virtualized networks. However, the potential price for enabling network slicing in multi-tenant/operator virtualized mobile networks is the underutilization of the scarce wireless and/or network resources. One way to increase overall network utilization would be to allow inter-tenant sharing, i.e., sharing of resources between different network slices. To this end, this thesis also aims to shed further light into this issue by discussing different possible degrees of network sharing together with the associated linear integer mathematical programs that allows to investigate upper bounds on the achievable performance improvement. Furthermore, in order to realize real time and adaptive sharing of resources a set of scale-free heuristic algorithms are also presented that is amenable for real-time implementation. Based on the depth of the aforementioned multi-tenant sharing, different sharing mechanisms are also specified in this thesis. In addition, by considering the advanced power controlling techniques in recent years, the proposed inter-tenant resource sharing framework is updated to a new level that allows variable transmit power during resource allocation, which potentially motivates higher resource utilization efficiency and energy efficiency in networks.

To present a comprehensive and practically applicable research, the last proposal of this thesis defines an optimization framework that enables optimal resource allocation with satisfying diverse service rate requirements, load offloading and efficient inter-tenant sharing in a practical fronthaul limited software computing based network. This proposal takes the realistic future network topology into account and mathematically defines the practical engineering constraints with the inter-tenant sharing scheme.

Based on these three main contributions in the thesis, some important issues related to 5G wireless architecturing and operation are explored and investigated in the perspectives of networks based optimization. Besides, the state of the arts in each proposal are all with their novelty, and the numerical experiments demonstrate the expected performance of all state of the arts.

1.2 Thesis Structure

This thesis consists of six chapters. Chapter 2 provides a comprehensive review on the background and some related existing approaches and concepts with respect to the current 5G wireless network development. The detailed explanation related to the cloud-RAN architecture, SDN, NFV and multi-tier multi-tenant network is provided. On the other hand, the recent popular state of the arts in network slicing, resource scheduling and resource allocation are discussed to provide a well-rounded comparison among these approaches. Finally, a brief introduction of Mixed Integer Linear Programming (MILP), the selected optimization model of technical works in the thesis, is also presented.

In Chapter 3, based on a conceptual Control Plane and User Plane Split (C/U split) architecture, a novel network functional decomposed topology and control plane signal optimization framework are proposed. The framework focuses on intelligently reduce the potential small cells handover rate and balance the control signal load in network. The problem is solved by using MILP solver in MATLAB and the results indicate a aggressive reduction in handover rate in considered topology compared to greedy method in present state of the art. Meanwhile, the load balancing is also achieved in the numerical results compared to greedy method. In addition, the potential trade-off between handover performance and load balancing performance is also investigated in this chapter.

In Chapter 4, by studying the present state of the art and principle of network architecturing, an optimization framework to flexibly share and reuse the OFDMA (Orthogonal Frequency Division Multiple Access) based physical resource block (PRB) among tenants in a virtualized wireless network is proposed. Moreover, to aggressively improve the overall network capacity and avoid overwhelming interference, the proposed framework is further combined with transmission power controlling technique and shed the light on different degrees of inter-tenant sharing. In technical aspect, this work defines a resource reuse maximization problem in a form of MILP, and provides two low complexity algorithms to achieve improvement in system throughput, per user rate and energy efficiency. To serve the principle of network architecturing, we introduce conceptual SDN architectures to implement the proposed framework in order to alleviate the chaos of controlling signals and inter-tier interference.

Chapter 5 brings the inter-tenant resource sharing to a more practical level in terms of wireless network engineering. In this chapter, a realistic wireless topology is considered which has capacity limited fronthaul linking the small cells to core network/cloud and various service rate traffic for users among different tenants. To tackle these strict practical constraints, a new inter-tenant resource sharing optimization framework is defined and linearized. The framework is tested in a set of numerical experiments and proposed optimal solution and algorithms show a satisfying gains in various network performance.

The conclusion is given in Chapter 6 and possible directions of future works are presented.

1.3 Contributions and Relevant Publications

The main contributions of this thesis are listed as follows:

- A conceptual control and user plane decomposed architecture is reviewed and the details of implementation the architecture is discussed. An optimization framework of control signal load balancing and handover rate reduction is proposed and performance of which is testified by numerical investigation.

- An optimization framework of engaging different degrees inter-tenant radio resource sharing is proposed and two sharing methods are defined. In addition, algorithms with different complexity are developed to solve the optimization problem efficiently and performance of which is testified.

- The inter-tenant resource sharing optimization framework is blended with variable power allocation techniques to further reduce network interference and sharing efficiency. Similarly, algorithms with different complexity are developed to solve the optimization problem and performance of which is testified.

- A practical emerging wireless network scenario is taken into account while applying the inter-tenant resource sharing scheme. A new optimization framework is developed to define this novel problem with constraints in fronthaul capacity and various user service rate. Regarding to this, an efficient algorithm called IRRA is developed to explore the sub-optimal solution of the proposal in polynomial time.

The relevant publications are listed as follows:

1. Zainab R. Zaidi, Vasilis Friderikos, Oluwakayode Onireti, Jinwei Gang, Muhammad A. Imran, "An Integrated Approach for Functional Decomposition of Future RAN", Energy Management in Wireless Cellular and Ad-hoc Networks, Springer International Publishing, 2016. (Chapter 3)

2. J. Gang and V. Friderikos, "Control plane load balancing in wireless C/U split architectures," 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, 2016, pp. 1-6. (Chapter 3)

3. J. Gang and V. Friderikos, "Optimal resource sharing in multi-tenant 5G networks," 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, 2018, pp. 1-6. (Chapter 4)

4. Jinwei Gang and Vasilis Friderikos, "Inter-Tenant Resource Sharing and Power Allocation in 5G Virtual networks", in IEEE Transaction on Vehicular Technology (Chapter 4, published, December, 2018)

5. Jinwei Gang and Vasilis Friderikos, "Service Aware Multi-Cell Transmission with Inter-Tenant Sharing in 5G Networks", in IEEE Transaction on Vehicular Technology (Chapter 5, accepted, January, 2019)

Chapter 2

BACKGROUND AND RELATED WORK

As already eluded in the previous Chapter, the focus point of the thesis revolve around the aspects of C/U split in 5G wireless network architectures and resource management operational issues. This chapter is organised as follows. Section 2.1 broadly introduces the baselines of designing the future wireless network. Section 2.2 details the software defined network (SDN) which is one of the most important techniques enabling 5G network functionalities and the optimization frameworks proposed in this thesis. Furthermore, Section 2.3 illustrates the fundamentals of network virtualization and slicing, in which different state of the arts and classic approaches related to virtual network slicing and resource allocation are introduced. In addition, Section 2.4 briefly introduces the mixed integer linear programming (MILP) in optimization problem which is the main modelling method adopted by the research in this thesis. At last, Section 2.5 concludes the information provided in this Chapter.

2.1 The Future of Wireless Network Evolution

Wireless communication network, initiated by the earliest voice-only communication system, has been rapidly improved and transformed towards a complicated, digitalized and intelligent system closely integrated with modern life of people. The gradually evolution of wireless network crosses from the first generation to the recent 4th Generation network (4G), which is stimulated by countless technological innovations such as digital modulations, internet based caching, efficient frequency reuse and etc. The current network, 4G, has proved that it is capable to improve the mobile phone voice quality, download/upload speed and service diversity compared to the previous generations of communication networks. However, some issues and challenges still rise to 4G network due to the exponential growth in mobile traffic with multi-media service requirement. For instance, a statistic provided by Cisco Visual Networking Index [1] saying that there will be more than half of connected smart devices appear in total communication traffic by 2019. In terms of smart devices traffic, larger size of download/upload traffic is expected. On the other hand, new applications and diverse service types including virtual reality, E-health related communications, high speed transportation communications, Internet of Things (IOTs) and Machine to Machine (M2M) communications [2] are also emerging that requires enormous data rate from wireless networks. Although 4G network is also advancing by applying new techniques such as Coordinated Multi-Point (CoMP) transmission, small cells and massive MIMO, however, without a conceptual and systematic evolution on wireless network itself the expected growing traffic in future is difficult to satisfy. Concluded by surveys in [3],[4],[5], the 5G wireless network shall be defined and identified as: 10 times data rates improvement compared to present 4G network, low round trip latency around 1 ms, supporting higher bandwidth for longer serving time in specific areas, connectivity to enormous number of devices (thousands for individual network area), practically available to users all the time, nearly 100% geographically coverage, reduction in electrical power by around 90% (green communications) and observable improved battery for network devices.

With the increasing amount of smart devices and communication applications, combining the exponential rise in service demand, a systematic transformation and update is urgently required by the existing cellular network. In terms of that, the following aspects are normally considered in both academia and industry [6] to build future 5G wireless networks: potential changes and innovation in architectural engineering related to RAN, breakthroughs in physical layer technologies (like mm-wave, directional antenna design and beamforming), and designing in MAC layer protocols and multiplexing schemes to support re-invented physical layer in cellular networks. Since the work in this thesis is mainly focusing on the first aspect, defining and solving optimization problem in horizon of realistic network architecturing, therefore the rest of two aspects will not be discussed in the thesis.

In terms of acrchitecturing, 5G wireless networks is expected to be a multi-tier (multi-layer) integration of all different size cells (macro, micro, pico and femto). Based on such extreme densification in network, space re-structuring is needed for network architecture design. As discussed in [6], a widely accepted solution is to deploy various size base stations in a HetNet structure which would improve radio frequency efficiency in 5G networks. Besides, the sectorized and multi-directional antennas of base stations is implemented because of the high co-interference while densely deploy different base stations in networks. In addition, to successfully deploy multi-tenancy in network infrastructure, network engineering also aims to fully apply SDN, cloud-RAN and network virtualization. Therefore, the following Section 2.2 and Section 2.3 respectively illustrates the details of heterogeneous topology adopted nowadays, SDN and Cloud based network (C-RAN) architecture, as well as network virtualization, NFV, spectrum slicing and resource allocation. More importantly, the current state of the arts related to these technologies are investigated in depth.

2.2 5G Wireless Network Architecturing Framework

2.2.1 Cells Densified Multi-Tier HetNets

In order to establish quality optimization frameworks based on future wireless network, a good understanding of HetNets, cells densifications, efficient network planning and deployment of small cells is necessary. In addition, this section reviews some classic works tackling interference management in multi-tier HetNets, which is a permanent consideration of the optimization problems in this thesis.

5G wireless networks need to provide communications between a large number of devices and diverse services, thus a proper integration of 5G cellular networks topology with the current LTE wireless networks is crucial. The main idea for such architectural transformation is to implement extreme cells densified multi-tier heterogeneous topology. A trend serving this idea is to move the base station centric topology towards a user/devices centric topology [7]. This topology redesign can be applied to both downlink and uplink transmission, as well as control and data forwarding signal links. Over the last decades, there are several original re-thinking ideologies based on the evolution from cell centric architecture to user centric architecture [7], which inspire the research works afterwards:

- Firstly, although the HetNet concept is originated from the 4G generation, however with denser cells deployment the coordination between transmit power and coverage areas of different cells is yet clear. From [8], a possible change on this could be implementing obvious different transmit powers and frequencies in different cells, decoupled the downlink and uplink signals, and distinct cell coverage. Authors in [8] specifies two different topology models (spatial model for HetNet base station deployment and Poisson model), to simulate the widely accepted network topology that contains enormous base stations with different sizes, transmit powers and coverages.

- The second proposal [9] illustrates an innovative way to support coexistence of multi-frequency bands in a single network infrastructure: using a so-called Phantom Cell, a specialized small size base station excluding control functions and working in high frequency, to separate the data forwarding functions from control plane. Based on this ideology, the control signal shall be sent by nodes transmitting high power and relative low frequencies (e.g. microwave frequencies) while the data signal is sent by nodes transmitting low power and relative high frequencies (e.g. mm-Wave).

- Thirdly, with the rise of network softwarization and centralization, the innovative concept of centralized baseband emerges with the rise of C-RAN [10] [11]. This concept aims to realize the decoupling between physical base station and the hardware that deals with the processing functionalities. The so-called Centralized Baseband Unit (central BBU) is responsible to virtualize the hardware functionalities and control them in a resources pool. This direction of thinking is a mainstream in 5G wireless architecture design, the details of C-RAN is further discussed in Section 2.2.2.

- Fourthly, in [7], architectural design ranged from fully centralized topology to fully distributed topology is discussed in details. Recall of such wide range of variation in topology design, the diverse service types (such as D2D, M2M and IoT) can be satisfied.

- Fifthly, the re-definition in Cooperative Multi-Point (CoMP) transmission/relaying system is also a direction to improve the network architecture in future [12]. Research from[13] suggests that the losses in half-duplex relays shall be recovered to some extents while performing CoMP transmission.

- At last, applying innovate network devices is also a promising way to improve the whole architecture performance [14]. With assistance from maturer IoT technologies, more and more smart devices can be implemented to the infrastructures, which definitely will positively impact the overall operation efficiency and energy efficiency in telecommunication networks.

Regarding to multi-tier HetNet topology, the deployment of small cells (including micro, pico and femto cells) is the key to architectural design. The goal of widely deploying large number of small cells in 5G networks is to efficiently compensate coverage holes and aggressively expand network capacity in communication areas [15] [16][17]. As discussed in [8], by adopting spatial base station model and Poisson model, densified cell deployment is applicable to a very complicated network topology. According to this, a rising concern is the complexity in interference control caused by complicated frequency reuse, especially when using OFDMA (Orthogonal frequency-division multiple access) among different types of base stations. Although the main interference in current networks design comes from inter-cell level rather than intra-cell due to the use of OFDMA, however the cell densified multi-tier HetNet topology itself brings multi-sources of inter-cell interference.

Based on this issue, authors in [18] propose a probability weighted based spectral resource allocation algorithm in multi-tier HetNet under C-RAN architecture. The main idea of this algorithm is to optimally decide if small base stations can use common spectral resources or dedicated spectral resources in networks based on a probability manner. The simulation results from which show improvement in networks frequency reuse while lowers users outage probability. In addition, the similar spectral resource assignment and interference control problems in two-tier HetNet are also studied by[19] [20] [21], however, these state of the arts have limitations such as that more complicated multi-tier (3 to 4 tiers) network topology is not applicable to the solutions, and the way of handling additional overhead signals generated from traffic offloading is not clear. As summarized in [22], the main challenges faced by 5G network researchers and network engineers in interference mitigations are the unbalanced geographical coverage and traffic load among different cells, restrictions in interference control caused by public or private access limitations and diversity of wireless devices (diverse transmit power and interference levels). These issues become more complicated while applying CoMP transmission or direct inter-user communications like D2D (Device to Device transmission). Beyond that, authors in [22] suggest a design guideline of distributed cell association and power control to tackle the interference issues in cell densified HetNet, where a combination of prioritized power control scheme and resource-aware cell association scheme is illustrated in details. As regarded to CoMP transmission, 3GPP release 11 [23] standardizes the coordinated scheduling scheme for multi-cells signalling and interference management which illustrates some core features of CoMP by work examples. However, such scheme is only based on certain CoMP devices, which in realistic network can be completely different from the 3GPP examples. With the same goal, researchers from Qualcomm Technologies Inc. and Samsung Mobile Solutions Lab [24] aim to tackle the interference issue by improve the performance of radio receivers to alleviate interference levels in present coding scheme, channel and resource allocation standards. Some proposals also focus on using power control techniques to efficiently address multi-tier cell interference in networks, and achieve energy efficiency at the same time. A green HetNets framework is proposed by [25], which specifically discusses the balance between spectral efficiency and energy efficiency in 5G multi-tier wireless networks. As summarized in [25], the optimal partition of the frequency bands used in networks with cell based power control is the fundamental to provide reliable interference control among different tiers. More works focused on the small cell deployment with variable transmit power, CoMP transmission in HetNet and cross tier interference control can be found in [26][27][28][29][30][31], which will not be explained in details due to the limitation of this thesis and the similarity between them and works mentioned above.

2.2.2 Developments within the SDN and C-RAN area

In terms of 5G network architecturing, another issue is to guarantee the intelligent coordination among different functional sectors in dense cell deployed networks. Based on this, concepts like Software Defined Network (SDN) and Cloud-Radio Access Network (C-RAN) have drawn massive attention of academia and industry. The smoothness control and cooperation provided by SDN and C-RAN are attractive, and based on these optimization frameworks and efficient algorithms can be designed. Furthermore, the backhauling and fronthauling issues are also introduced in this section, which is a key consideration in Chapter 5.

Regarding to the definition, different organizations (e.g. Open Networking Foundation (ONF) or Software Defined Networking Research Group (SDNRG)) have slightly different but generally common standard to recognize a SDN. Generally speaking, the SDN is an architectural framework that provides programmable networks functions [32][33] with separation of the network control plane from the data forwarding plane. The programmability, centralized management and adaptability provided by SDN can ensure that high bandwidth intensive applications like video streaming will be fully supported. The most innovative part of SDN is the proposal of physical/logical decoupling of user plane (data transmissions) and control signalling paradigm [34][35][36][37], popularly recognized as C/U split. Imagine a C/U split SDN architecture with extreme intense cell deployment, the macro BSs [1] provide wide area coverage operate as control (signalling) nodes and are responsible for the overall orchestration and resource management of a large number of small cells within the coverage area and how mobile users are connected to them [38]. By introducing such an independent control plane in architecture, simplicity, swiftness and agility of control functions can be achieved in cell densified multi-tier networks [39][40][41]. In the horizon of overall network design, the main advantages of C/U split architecture can be concluded as followings:

- High data rate can be achieved in certain areas by freeing data forwarding focused nodes from control functionalities.
- The control overhead signal can be simplified due to the centralized control plane in networks [35].

[1] for simplicity, base station/cell is expressed exchangeable as BS henceforward

- Hardware constraints in different cells can be efficiently reduced due to the powerful software components embedded in centralized control plane [42].
- Easier interaction between control plane and data plane can be realized by using solid interfaces [43] [44].

Speaking of physical design, the macro BSs include, inter alia, the control plane interface with the Evolved Packet Core (EPC) entities such as the Mobility Management Entity (MME) via the 3GPP standardized S1-MME interface. EPC is a core network architecture introduced in LTE networks, which consists of several network elements handling traffic for base stations [44]. In addition to that, C-plane will also include all LTE signalling and control related functionalities in some cases, such as the radio resource control (Radio Resource Control (RRC) - establishment, modification, and release of mobile users RRC layers) network controlled mobility and functionalities related to measurement, configuration and reporting [45][46]. To ensure full coverage it is also possible for macro BS in addition to control signalling only (C) mode of operation to provide low bit rate services to mobile users (provides functionalities as C+U). On the other hand, small cell nodes would be activated 'on demand' and will deliver ultra high bit rates under a small footage area. This term of operation is described as dual connectivity within the 3GPP, since it allows mobile users to maintain simultaneously C-plane and U-plane connectivity from macro and small cell respectively[47] [48]. Summarized in [49] and [33], the data plane normally contains network elements such as virtual switches and physical switches. By accessing these switches, data plane can smoothly collect information from control plane and forward data/packets to end users. On the other hand, the control plane is totally software based and is responsible for integrating all the RAN elements in networks and managing radio resources.

Similar to SDN, Cloud-Radio Access Network (C-RAN) concept, proposed by China Mobile [50], has also attracted much attention from academia and industry [51][52]. Unlike the traditional networks, where BSs have integrated processing unit and radio heads, C-RAN can provide efficient coordination among processing capacities of BSs especially when network virtualization is applied [53][16]. Network virtualization originally refers to a process that combining hardware, software, network resources and network functionalities to a single centralized entity, called virtual network, which will be detailed in Section 2.3. In addition, networks energy efficiency and operation capitial reduction can be achieved due to the decoupling Baseband Unit (BBU) in a C-RAN architecture.

In general, a typical C-RAN is recognized when three key features appear: centralized BBU, distributed Remote Radio Heads (RRHs) and fronthaul links [53][54][55][56]. The centralized BBU, recognized as a set of physical servers in data center or control plane equipped with real-time virtualization technologies and fast processors, combines all the computational radio resources among base stations in networks. It functionally manages and assigns those resources (virtual resources in virtualization) as a resource pool. In this way, simpler communication and coordination between BSs can be realized with the reduced energy consumption and radio devices cost because of the processing unit removal in RRHs. Speaking of the RRHs, they collect information and virtual resources from BBU then forward extreme high data to the nearby users as the data planes in SDN. To guarantee the seamless communication between BBU and RRHs, the role of fronthaul links is another key for a well designed C-RAN architecture. The main requirements of fronthaul connections in C-RAN high bandwidth, low latency and tight synchronization. The traditional fronthaul cannot satisfy these requirements when massive RRHs are deployed and extremely high data traffic exist [57][58]. Therefore, some potential solutions are investigated for the recent years such as the Millimeter-Wave Multihop Fronthaul [59][60] and Free Space Optical fronthaul [61][62]. In addition, nowadays all the fronthaul engineering follows the Common Public Radio Interface (CPRI) [63], a commonly used standard. The CPRI states that the capacity of fronthaul design normally varies numerically based on the cell deployment density, geographical environment and traffic scenarios. Regarding to he high cell intensity and complex traffic scenarios in future 5G networks, the remain barriers for fronthaul engineering can still be the high cost and capacity limitation. Such issues are considered by technical work following the CPRI standards in Chapter 5. Besides, the difference between fronthaul and backhaul in terms of a ultra densified wireless network shall be clarified. Unlike fronthaul connecting the RRHs to the BBU based control center, the backhaul is defined as the physical link between base stations or BBU center[64], and the core networks thus normally backhaul capacity is much larger compared to the one of fronthaul[65][66][67][68][69].

In summary, the SDN and C-RAN are both popular concepts to bring extremely high user data rates, expanded network capacity, easier inter-cell cooperation, energy efficiency, efficient radio resource management and expenditure reduction to future wireless networks[70]. Based on SDN and C-RAN, more network proposals are developed in recent years, some examples including Wireless Network Cloud (WNC) [71], Light Radio[72] and Super Base Station architecture[73]. Although these networks have various design considerations, they all have similar principles such as enabling C/U split and centralizing the resources and control functionalities.

2.3 Wireless Network Virtualization and Resource Allocation

With the tremendous growth in mobile network traffic and the need to accommodate new vertical markets in the mobile/wireless ecosystem, a proper coexistence and isolation of multi-mobile carrier networks in the same physical network infrastructure becomes a new requirement to 5G. In this case, wireless network virtualization to RAN is recognized as one of the key technologies [74]. By applying virtualization technology, a physical network can be sliced into multiple virtual networks meanwhile the functionalities in networks are also decoupled by Network Function Virtualization (NFV) [75][76][49]. In terms of wireless resources, network slicing is a more specific concept to realize a virtual network, which is usually defined as the assignment of a subset of network resources or functionalities to certain tenants for supporting their end users [77]. Although wireless network virtualization requires a large room for improvement, it is still a promising tool in realizing a complicated and applicable 5G wireless network especially with the assistance from SDN and C-RAN[74]. The technical works in this thesis propose different optimization frameworks which are built on virtualized networks environment. Therefore, a comprehensive review of network virtualization, NFV, network slicing and virtual resource allocation problems is hereby presented.

As mentioned at the beginning of this chapter, the expected exponential growth in mobile traffic and extreme diverse service types can be a tough mission to Mobile Network Operators (MNO) or Service Providers (SPs)[2]. To service providers, the high network management complexity and limited radio resources means a

[2] In this thesis the terms MNO, tenant and SP are used interchangeably.

necessary increment in both capital expenses (CAPEX) and operating expenses (OPEX). Fortunately, the emerging trend of utilizing wireless network virtualization, SDN and C-RAN offers them a promising option [78][79].

The principle of network virtualization is to orchestrate and extract physical network resources to virtual resources that can be shared by all parties who need access to these resources without actually owning it [80]. The advantages brought by network virtualization include flexible network resource utilization, reliable system performance, improved service rate at end users, reduced CAPEX (Capital Expenditure) and OPEX (Operational Expenditure), as well as simplicity in re-configured while applying new technologies. Once virtualization performed, the infrastructure provider (InP) and SPs will share the physical network. In this case, InP is still the one that owns the infrastructure and radio resources in reality, and SPs are tenants that lease the network, also named Mobile Virtual Network Owner (MVNO). Furthermore, the SPs leases the virtual radio resources from the virtualized networks then assign the resources to their end users. Regarding to the virtual resources, they are abstracted to various slices belong to different InPs/tenants, meanwhile all the other network elements are virtualized and included in the slices.

Summarized by [74], there are three different levels of slicing within network infrastructure: Spectrum level, network level and flow level. In spectrum level, the radio resources are ultimate targets for slicing [81]. This process actually is similar to a wide range of dynamic resource partition and reuse. The network controller in this case is in charge of scheduling network resources to different slices, base stations, and users. The spectrum level slicing is the main consideration in this thesis since this thesis solves different resource reuse and sharing problem based on spectrum slicing. In the eye of telecommunication business, nowadays the spectrum in a shared network can be divided to exclusive one and unlicensed one. The exclusive spectrum cannot be shared thus interference control is simpler [82][83]. On the other hand, unlicensed spectrum are open for access which means they can be reuse by any parties registered in the networks thus higher interference is expected. Therefore, intelligent spectrum sharing and resource allocation approaches are important to solve this problem[84][85][86]. Apart from that, the network level slicing indicates the virtual slices containing some physical elements such as BSs, service gateway, CPU and etc, which are usually made based on resource requirements, link quality, packet budget, and transmit power. The flow level slicing is normally referred to a slice based aggregated data rate or bandwidth. The main aim of that is to lease certain bandwidths or rates to SPs that have users to serve but do not have any resources in their own physical networks.

2.3.1 Network Function Virtualization (NFV)

In a virtualized wireless network, the role of Network Function Virtualization (NFV) is to enable a fully computational environment. The original idea of NFV is to run all network functions from hardware appliances on cloud computing infrastructures when network virtualization is applied. The details of NFV implementation (e.g. what type of components considered in a virtualized platform) can be found in [87], which is not the focus of this thesis. To be noted, running virtualized functions in cloud infrastructure does not make NFV equivalent to a commercial cloud. However, the emerging of more advanced computing technologies can improve the performance of NFV.

The benefits brought by NFV include massive hardware reduction in network systems, lower power consumption on devices, lower infrastructure cost, and simplicity in network upgrades. From the perspectives of network architecturing, NFV assists SDN by providing management/cooperation within all network entities without hardware restrictions as in traditional networks. Besides, NFV is also important to C-RAN as the virtualized functionalities of all entities can be stored in the cloud or central BBU that is responsible to assign all these functionalities when there is relevant demand in networks [87]. A popular example for this is the NFV utilization in Evolved Packet Core (EPC), a core network with different functionalities entities in LTE network [88][89].

2.3.2 Resource Allocation in Virtual Networks

Spectrum scarcity and radio resource allocation are always seemed as research challenges to the development of 5G wireless networks. To tackle the spectrum shortage issues, some proposals such as spectrum aggregation[90] and utilization of mmWave frequencies [91][92] have drawn massive attentions as a promising solution to expand spectrum capacity. However, the optimal way to assign the radio resources from the expanded spectrum is still the concern for future networks. Therefore this section reviews a wide range of resource allocation solutions in both traditional networks and 5G targeted virtualized networks.

There are several conventional techniques realizing efficient resource allocation in wireless networks [93].

- Traffic application-aware based allocation: This type of resource allocation approaches aim to use logarithmic and sigmoidal-like utility functions to represent the applications of users. Some optimization related to that can be found in [94] [95]. However, the main weakness is that there is no distinction among users in different tiers that some users might have higher priority in QoS.

- Simple spectrum matching theory based allocation: as stated in [96], another approach of resource allocation can be simply adopting distributed matching theory, which is a low complex algorithm. In this theory, the allocation scheme is completely depends on stable matching between different assignment elements such as available radio resources and transmitters [97].

- Spectrum auctioning based allocation: this approach sees the users as bidders competing for available spectrum resources with different biding price, and networks nodes as agents having a cost for the resources. The bidding process ends while all resources are assigned. Zhang et al. present a detailed review on different auctioning methods in wireless system [98].

- Priority based allocation: as discussed in [22] , the traffic based and tier-based allocation are the mainstreams in this type of approach. Traffic based allocation assign resources based on different QoS requirements of users, whereas tier-based allocation assign resources based on priority of different networks tiers such as macro BSs, pico BSs and femto BSs.

However, in a virtualized wireless network, there is an add-on issue to all these classic approaches: resource sharing. In a network with resource sharing, different parties or tenants would require the network to share its radio resources, functional resources and management resources in different levels, to efficiently serve their own users which normally have various bandwidth and traffic needs. Therefore, facing such a challenge,interference management, efficient resource reuse and scheduling among different virtual slices for multi-tenants would become more complicated. Based on the present state of the arts, solutions dedicated to perform intelligent slicing and virtual resources allocation are widely studied. The authors in [76] give a detailed comparison between different sharing approaches that could be applied to general or specialized wireless architecture, which distinguish resource sharing in capacity, spectrum and base station level, respectively. However, the main works can be summarized as the followings.

The most common and widely used approach for resource slicing is called static or fixed resource slicing [99]. The idea of static slicing is simply reserve a fixed portion of resources for each network tenant and this reservation cannot be changed afterwards. However, the static slicing might lack of efficiency when the resource is limited and traffic episode has tremendous variation. Therefore, another approach named dynamic slicing is proposed as a popular state of the art [100][101] and the main idea of which is to instantly redistribute network resources to different tenants based on their temporal traffic and resource requirements, by utilizing certain advanced orchestration functionalities in SDN or C-RAN. In fact, before dynamic slicing, in order to motivate flexibility in traditional static resource reservation (SR)/fixed slicing, a well-known concept named Network Virtualization Substrate (NVS) is proposed and enhanced in [102][103][104], which specifies an association of two resource sharing approaches with customized scheduling techniques. The two approaches are named bandwidth based NVS and resource based NVS, indicating resources slicing based on application/service bandwidth requirement or amount of aggregated resources required by a tenant respectively. Unfortunately, the NVS approaches only consider single cell scenario therefore in a multi-tier dense network the ability of NVS will be limited. Considering multi-tier network with dense cell deployment, work in [105] presents an optimization framework to jointly allocate power and subchannel resources in a two-tier network with virtualization with consideration of mobility. Similarly, a 5G network based control plane is proposed by [106] to perform trade-off between resource slicing and user mobility. On the other hand, works such as [107][108] intend to apply game theory to resource slicing optimization that dynamically assigns system bandwidth as auction games in virtualized wireless networks. However, the game theory based approaches still face difficulties to cope with issues like underutilization of resources and infeasibility to re-assign resources when traffic dynamically varies. Thus, further research works which are similar to dynamic slicing have been proposed. By re-assigning radio resources between tenants in a short time period, authors in [109][110] demonstrate that their improved dynamic slicing approach can compensate certain degrees of resource underutilization in future networks. Meanwhile authors in [104] also propose a novel version of NVS with so called PRR (partial resource reservation) technique to reserve certain 'mutual resources' in networks that can cope with dynamic traffic in some extents.

2.4 Mixed Integer Linear Programming Model

The optimization modelling completed in this thesis is based on the Mixed-Integer Linear Program (MILP), which is modified from the classic Linear Program (LP). In MILP, the variables in model can only be 0 or 1. An example of using MILP in a simple wireless networks problem can be deciding if a user can be connected to a base station, if connect the variable is 1 otherwise is 0. With certain constraints (rephrasing the logical requirements of a problem) that restrict the combination of variables, the relevant problem solution can be calculated based on the discrete decisions in model [111][112]. Demonstrated by [113], the MILP problem normally is NP-complete or NP-hard in complexity. A solution to MILP is usually hard to obtained. The method normally used to accelerate the solution computation is called branch and bound algorithm [111]. The main idea of branch and bound is to explore subsets of the solution set as branches, meanwhile exam the branch according to the estimated lower and upper bounds of the optimal solution found so far. More details for branch and bound can be found in [111]. Since the technical works in this thesis is evaluated based on software MATLAB, the MILP solver is used (a well programmed optimization tool) to compute the solutions to all the optimization problems proposed in the thesis [114].

However, most of the technical works in this thesis involve thousands of variables, which even makes the computation of MILP solver impossible. Therefore, heuristic algorithms are always designed to provide solutions with certain compromise in performance depending on the nature of each optimization problem,

2.5 Conclusions

In this chapter, the relevant state of the arts and literatures are reviewed to provide a thorough understanding of future wireless networks and research gaps. In Section 2.1, a brief introduction on the development and current challenges of 5G wireless networks is presented in a broad horizon. Section 2.2 provides a detailed background review on 5G networks from the perspective of network engineering, where some popular 5G network planning concepts and topologies such as cell densified HetNet, SDN and C-RAN are illustrated in details. In Section 2.3, network virtualization is introduced, which is deemed as one of the key techniques to achieve performance of 5G networks. In details, this section illustrates the origin of network virtualization, utilization of NFV, and the widely researched resource slicing/allocation problem in virtual networks environment. At last, Section 2.4 briefly introduces the concept of linear programming, which constructs the fundamental of mathematical modelling in technical works presented in this thesis.

Some documents are cited in this section for readers with interests in understanding linear programming modelling in a better sense.

From these literature reviews, some research gaps have been discovered, which translated to research works proposed in this thesis. From the best knowledge of thesis' author, a gap found in current C/U split network topology proposals is that there is no methodology considering management of control plane signals, which could become overwhelming while all control traffics are forced to a single control plane in a massive network. Other gaps found in present literatures are mainly related to multi-tenants resource slicing problems. Although there are some current state of the arts that aim to assign resources optimally in a shared network (such as NVS and game-theory based resource slicing), there is no methodology contributes in providing flexible resource sharing between tenants in a dynamic way. This inspires the research in Chapter 4. In addition, with understanding of engineering technique like joint transmission, the resource sharing proposal are enriched to another level, which improves networks resource efficiency further, in Chapter 5. In order to ensure the quality of these works, the 5G background concepts and standards are frequently reviewed and applied to the design of research. For example, all of these proposals are well embedded in the current SDN or C-RAN based architectures with C/U functional split, instead of a random network architecture with endless assumptions. Besides, numerical experiments in these proposals are also following different 3GPP releases, and present widely cited academic papers.

Chapter 3

CONTROL PLANE SIGNAL LOAD OPTIMIZATION

The goal of expected 5G networks is to bring ultra high data rates to mobile users. To realize this, a novel Control and User plane split (C/U) communication network paradigm that deploys a large number of small base stations within the coverage area of a macro cell is been considered. However, an emerging problem for such architecture is the increasing complexity in control network load balancing and hand over events. In academia, load balancing for the control has received little attention and it is the focus of this paper thus a set of optimal solutions to the problem of load balancing for multi macro cell controllers are proposed and testified via numerical investigations

3.1 Introduction

Undoubtedly, cell densification in 5G wireless networks is expected to allow for ultra high achievable data rates. In order to efficiently realize high cell density networks, an architectural paradigm shift is taking place towards the physical and/or logical split of the Control and User planes (C/U planes). Since user plane performance is inevitably intertwined with that of the control plane, special attention should be paid for cases of congestion episodes, where limited control plane capacity might adversely affect overall network performance.

In this chapter, firstly, the SDN based control plane is investigated in details because the traffic flows, both controlling signals and hand over signals, managed by it are deemed as the targets to be optimized in the proposed optimization problem. In other word, SDN central controller is the platform where the optimization carries out. Secondly, a network optimization framework for C/U split SDN architectures with emphasis on control plane load balancing and C-plane load reduction by taking into account inter macro controller handover and back-hauling limitations is defined and tested. Numerical results reveal that proposed optimal solutions provide significant gains in terms of network performance compare to baseline technique.

The control and user (i.e. data forwarding) plane separation (C/U split) is expected to provide a significant facilitation towards the co-existence of different radio technologies including high capacity small cells within the same network that are orchestrated by a common control infrastructure. Such re-thinking on the cellular architecture is required to accommodate future needs; it is now estimated that by 2020, mobile user demand for data will generate an order of thousand times the traffic level on mobile networks compared to year 2010. In a C/U split plane network domain macro cells act as the network control plane that efficiently manage real time resources of a large number of high capacity small cells (including future mmWave-based cells) that can serve mobile users always ‘on demand’. To handle the increasing data demand from mobile users it comes as no surprise the move towards the use of very dense, low-power, small cell wireless networks that will unlock the potential of extremely high spatial reuse (especially when utilizing spectrum at the mmWave bands). Small cell networks allow for increased capacity by spatial localized transmissions (i.e. bring small access points closer to the user) whilst trading off a more efficient spectrum utilization by allowing multiple concurrent connections to different access points for a mobile user. The concept of C/U split has been envisioned in the setting of small 'phantom' cells [1] that will serve mobile users (U-plane) while being controlled by a macro base station operating at different frequency bands [115].

Hand in hand with the aforementioned benefits arrive a set of challenges when moving towards the envisioned extreme cell densification scenarios [116]. Although the actual percentage of control signal among all type of networks signal is unknown, the complexity of a multi-cells multi-layers network in future would require a dramatic growth in control traffic. On the other hand, a survey [117] shows the potential handover event consumes a large part of transmission signals in current LTE network standards, which can be up to 40% of total signals. Not least among those, and items of attention in this study, are issues related to macro base station and small cell association and coordinated management that will become an increasingly difficult and complex network function in the near future. Therefore, the potential significant signals for controlling and handover can pull back the overall network performance in 5G. To this end, a detail optimization framework aiming to increase the performance of C/U split architectures under the assumption of a potentially large number of small cell and associated user mobility. More specifically, more interest has been paid to dynamic policies for assigning small cells to macro base stations in the overlapped control area (controlled by

[1] the terms pico cells, small cells and phantom cells are used interchangeably in this thesis.

multiple macros BS/control planes) by taking into account control plane load conditions aiming to avoid degradation on the performance due to congestion episodes. Besides that, seeking a way to reduce as much handover rates as possible is also in consideration since the load re-distribution would easily cause massive handover in a architecture with dense small cells. In a essence, SDN enabled wireless access and core networks provide a hardware agnostic programmable framework for easing the development of new network functionalities and isolating complexities through the separation of control and data plane. SDN will allow the required flexibility in network monitoring, policy installation and network management and hence act as a catalyst in allowing novel network orchestration techniques to be adopted in the C/U plane split wireless architectures as envisioned in this chapter.

3.2 Control Plane in Future Networks

As discussed in Chapter 2, the control plane in 5G wireless networks is a core element to provide key functionalities in operations. It is necessary to break down the concept of control plane in current network standards to illustrate how it will be integrated with the optimization framework. An example of SDN based control plane is illustrated in Fig.3.1 [118]. Although different SDN based architectures have slightly different design of control plane, however, control plane normally can be located in a data center of core network/central cloud, the cloud edge or MBS. If control plane is implemented in a MBS , such MBS is able to manage the whole system resources horizontally and guide SBS or RRH to assign virtual resources to users without cross tier interference. To support SBS, control signalling for managing the Phantom Node shall be sent by MBS via certain “New Interfaces.”

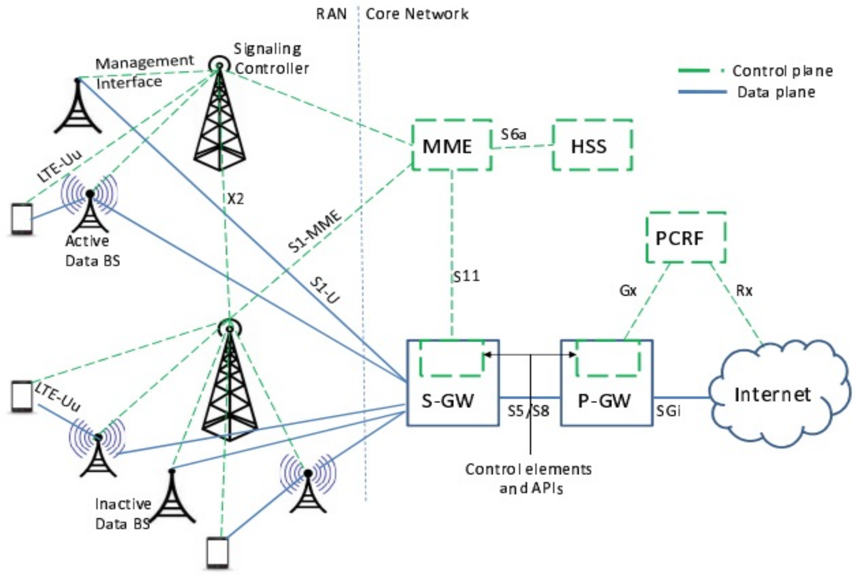


Fig. 3.1 The SDN based controller should contain some functional capabilities currently residing in S-GW (Serving Gateway) and P-GW (Packet Gateway) along with the dedicated control nodes of EPC (Evolved Packet Core) shown with green dashed lines in Fig. 1. The functional decomposition of S-GW and P-GW is shown by means of control elements containing necessary APIs to EPC control nodes and also control logic required for data forwarding. These programmable control elements for routing modification etc., can be housed in control nodes in future architectures with specific APIs only in S-GW and P-GW

In most configuration, the relationship between the MBS and the SBS is master-slave rather than peer-to-peer [36]. The procedure of connecting a SBS to a UE such as demodulation, channel estimation and measurements, and Cell Reference signalling (CRS) is initiated and completed by control plane. Due to the control plane and SBS using different frequencies, the UE shall perform inter-frequency detection to discover both of them. In addition, the control plane is responsible for synchronizing discover signal thus the UE can detect the signal faster.

The link for establishing the new interface between MBS/control plane and SBSs is fronthaul. The capacity of which is also another issue in network engineering. On the other hand, each MBS or control plane is also linked to core networks by backhaul, although backhaul has larger capacity compared to fronthaul, the efficient use of such capacity is still a important consideration for operating the control plane, especially when MBS provides both control and data signal in some scenarios. According to this, functional optimization related to the control plane signal load is necessary.

3.3 System Model and Problem Formulation

In this section, network scenario is defined and formulated to establish the related optimization problems for performing load balancing and inter macro cell handover minimization in order to improve the performance of a generic C/U plane split wireless architecture.

3.3.1 Scenario and Problem Formulation

As eluded previously this work assumes a future high dense network scenario with a cell density of at least ten times higher compared to current levels as expected in 5G networks and in addition to the pico cells we also assume a number of macro-controllers that have overlapping coverage areas. The proposed set of optimization problem relate to the cases where a pico cell can be served by a number of different macro controllers (See Fig 3.2). This is a pragmatic assumption since such information is readily available from mobile network operators. This framework assumes a set of pico cells and macro controllers P and M respectively. To proceed with a mathematical programming setting we define a binary decision variable x_{ik} as shown in Eq (3.1).

$$x_{ik} = \begin{cases} 1, \text{ if pico cell } i \text{ controlled by macro } k. \\ 0, \text{ otherwise.} \end{cases} \tag{3.1}$$

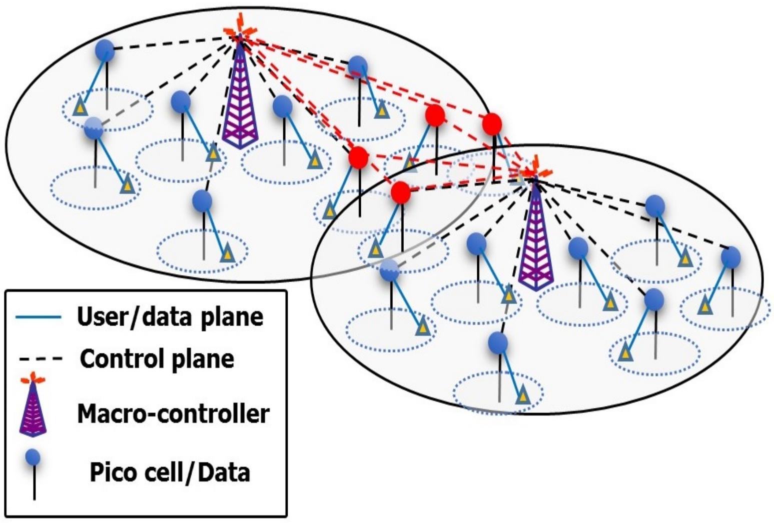


Fig. 3.2 A generic C/U split wireless architecture with a number of pico cells (each serving 1 user) and red ones are those could be controlled by both macro controllers.

In addition, the framework also defines the average handover rate from pico cell i to neighbor pico cell j by h_{ij}. Furthermore this framework assumes that each macro controller k has a capacity of C_k (Mbps), which limits the upper-bound of controlling signal for a control plane, and each pico cell i based on its current load has a required channel for control signal from control plane as g_i (Mbps or Kbps), which captures the required L1/L2/L3 control signalling flow including overhead for admission control and synchronization. As mentioned, information about average handover rates is readily available using historical data.

3.3.2 Problem Formulation for C-Plane Overhead Reduction

Based on the above network setting we can now define an integer mathematical program for C-plane overhead signalling reduction. The following optimization problem focus on reducing C-plane overhead by assigning pico cells to macro controllers using, previously known, average handover rates between different pico cells. More specifically, a non-linear integer mathematical optimization problem that reduces the overhead on C-plane can be defined. This framework calls this the optimal handover (OPT-HO) problem and defines it mathematically as,

$$[\textbf{Prob 1}] \ \min \sum_{k\in\mathbf{M}} \sum_{i\in\mathbf{P}} \sum_{\substack{j\in\mathbf{P},\\ j\neq i}} h_{ij}(1 - x_{ik}x_{jk}) \tag{3.2}$$

s.t.

$$\sum_{i\in P} g_i x_{ik} \leq C_k, \ \ \forall k \in \mathbf{M} \tag{3.3a}$$

$$\sum_{k\in M} x_{ik} = 1, \ \ \forall i \in \mathbf{P} \tag{3.3b}$$

$$x_{ik} \in \{0,1\}, \ \ \forall i \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.3c}$$

The objective function (3.2) shows that the overhead costs will be accumulated if handover happens between two pico cells which are controlled by different macro controller. Also note that the objective function is quadratic and hence non-linear, therefore linear programming techniques cannot be applied in this form. Constraint (3.3a) ensures that the capacity of the macro controller in terms of active supported flows is not violated. Constraint (3.3b) enforces that each pico cell can only be connected to one macro-controller and constraint (3.3c) denotes that the decision variables are binary.

However, as already mentioned, the objective function is not linear which deems the above Prob 1 formulation not suitable for utilizing powerful integer linear programming solvers. To re-formulate the above problem into an integer linear problem we introduce the following integer decision variable $z_{ijk} = x_{ik}x_{jk}$ with $i, j \in P$ and $k \in M$. As can be seen the variable z_{ijk} takes the value of one only when pico cells i and j are connected to the same macro controller k, otherwise it takes the value of zero. More specifically variable z_{ijk} can be formally defined as follows,

$$z_{ijk} = \begin{cases} 1, \text{ if pico cell } i, j \text{ controlled by macro } k. \\ 0, \text{ otherwise.} \end{cases} \tag{3.4}$$

With that in mind, the problem can be re-formulated in an integer linear mathematical program as follows,

$$[\textbf{Prob 1'}] \ \min \sum_{k \in \mathbf{M}} \sum_{i \in \mathbf{P}} \sum_{\substack{j \in \mathbf{P}, \\ j \neq i}} h_{ij}(1 - z_{ijk}) \tag{3.5}$$

s.t.

$$\sum_{i \in P} g_i x_{ik} \leq C_k, \forall k \in \mathbf{M} \tag{3.6a}$$

$$\sum_{k \in M} x_{ik} = 1, \ \forall i \in \mathbf{P} \tag{3.6b}$$

$$z_{ijk} \geq \ x_{ik} + x_{jk} - 1, \ \forall i, j \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.6c}$$

$$z_{ijk} \leq x_{ik} \quad \forall i, j \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.6d}$$

$$z_{ijk} \leq x_{jk} \quad \forall i, j \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.6e}$$

$$z_{ijk} \in \{0, 1\}, \ \forall i, j \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.6f}$$

$$x_{ik} \in \{0, 1\}, \ \forall i \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.6g}$$

Constraints (3.6c)-(3.6e) are used to bind two variables in optimization problem.

3.3.3 Load Balancing Problem Formulation

In addition to minimize the overall control overhead the framework could also provides an association of pico cells to macro controllers so that to achieve load balancing across the various macro controllers. Based on the previous definitions the control related load F_k that has to be handled by macro controller k can be written as follows,

$$F_k = \sum_{i \in \mathbf{P}} g_i x_{ik} \tag{3.7}$$

Providing load balancing between the macro controllers can be deemed as an important requirement in order to allow for efficient utilization of scarce wireless resources. To this end, a macro controller load balancing problem can be defined. The optimal load balancing (OPT-LB) is defined as follows,

$$[\textbf{Prob 2}] \;\; \min \sum_{k \in \mathbf{M}} \left[\sum_{i \in \mathbf{P}} g_i x_{ik} \right]^2 \tag{3.8}$$

s.t.

$$\sum_{i \in P} g_i x_{ik} \leq C_k, \forall k \in \mathbf{M} \tag{3.9a}$$

$$\sum_{k \in M} x_{ik} = 1, \;\; \forall i \in \mathbf{P} \tag{3.9b}$$

$$x_{ik} \in \{0, 1\}, \;\; \forall i \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.9c}$$

The objective function in Prob 2 is defined as load balancing indicator which indicates a more balanced load scenario while it has smaller value. As shown previously, the proposed optimization function with constraints are noted as a non-linear integer mathematical programming problem which needs to be linearized in order to utilize mixed integer linear programming solvers in MATLAB. Therefore, the above defined Prob 2 can be re-formulated as a linear problem if viewed as a max-min optimization problem. In this case, the problem can be re-formulated as follows,

$$[\textbf{Prob 2'}] \ \max \ t \tag{3.10}$$

s.t.

$$t \leq \sum_{i \in P} g_i x_{ik}, \forall k \in \mathbf{M} \tag{3.11a}$$

$$\sum_{i \in P} g_i x_{ik} \leq C_k, \forall k \in \mathbf{M} \tag{3.11b}$$

$$\sum_{k \in M} x_{ik} = 1, \ \forall i \in \mathbf{P} \tag{3.11c}$$

$$x_{ik} \in \{0,1\}, \ \forall i \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.11d}$$

3.3.4 Joint Problem Formulation

Without loss of generality, the load balancing and handover reduction problem shall be equally important in network engineering. However, it is also interesting to investigate the joint performance of these two under certain requests, hence the objective functions of Problems 1 and 2 could also be considered jointly by creating a weighted sum objective function to find trade-off between load allocation across the macro controllers and reduction of the C-plane overhead. Such a joint optimization problem can be defined as follows using scalar weights ω_1 and ω_2 to define the contribution in the objective function.

Before formally present the joint performance objective function, the normalization of variables shall be conducted first. Two new function $f_1(z_{ijk})$ and $f_2(t)$ are defined based on the physical meaning of Prob 1' and Prob 2'.

$$f_1(z_{ijk}) = 1 - \frac{\sum_{i\in\mathbf{P}}\sum_{j\in\mathbf{P}}\sum_{k\in\mathbf{M}} h_{ij}z_{ijk}}{\sum_{i\in\mathbf{P}}\sum_{j\in\mathbf{P}} h_{ij}} \in [0,1] \tag{3.12}$$

where $h_{ij} = 0$ when $i = j$.

$$f_2(t) = \frac{t}{C_k} \in [0,1] \tag{3.13}$$

Prob 1' indicates the cumulative potential handover rates in control signal which has the maximum value as the summation of h_{ij}. Therefore, function $f_1(z_{ijk})$ embeds Prob 1' with its maximum value that brings handover rate optimization to 0 to 1 scale. Similarly, function $f_2(t)$ combines Prob 2' with the potential maximum value of a macro cell control load C_k, which scales the load balancing optimization output to 0 to 1 range.

Based on all the transformations, the joint optimization function is defined as,

$$[\textbf{Prob 3}] \min\{\omega_1 f_1(z_{ijk}) - \omega_1 f_2(t)\} \tag{3.14}$$

s.t.

$$\sum_{i\in P} g_i x_{ik} \le C_k, \forall k \in \mathbf{M} \tag{3.15a}$$

$$\sum_{k\in M} x_{ik} = 1, \ \forall i \in \mathbf{P} \tag{3.15b}$$

$$t \le \sum_{i\in P} g_i x_{ik}, \forall k \in \mathbf{M} \tag{3.15c}$$

$$z_{ijk} \geq\ x_{ik} + x_{jk} - 1,\ \ \forall i, j \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.15d}$$

$$z_{ijk} \leq x_{ik} \ \ \forall i, j \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.15e}$$

$$z_{ijk} \leq x_{jk} \ \ \forall i, j \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.15f}$$

$$z_{ijk} \in \{0, 1\},\ \ \forall i, j \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.15g}$$

$$x_{ik} \in \{0, 1\},\ \ \forall i \in \mathbf{P}, \forall k \in \mathbf{M} \tag{3.15h}$$

The objective for this joint optimization problem is to minimize the overall handover traffics ($f1$) and maximize the control load in a least congested macro controller ($f2$) at the same time. Therefore, to use a single minimization sign for the whole function, second objective is subtracted by the first one.

It is clear that ω_1 and ω_2 are weight factors for handover reduction part and load balancing part respectively, and ranged between 0 to 1. Except the objective function, all the constraints from Prob 1' and Prob 2' are remained.

3.4 Comparison Method: Distance Based Cell Allocation

A nominal mechanism can be assumed, in order to compare with the proposed set of solutions, when a pico cell to macro controller allocation takes place via the Euclidean distance between them or the mean link channel conditions. In this case, pico cells are allocated to a macro controllers which is closest to them compared to proposed scenario that always allocates pico cells pairs with high handover rates to the same macro. Such greedy based allocation can result in significant sub-optimal solution and we provide below a theoretical result on the sub-optimality of such greedy allocation compared to our optimal solution. As shown below, using such a metric the optimality gap is in essence unbounded.

Lemma 1: The gap between the optimal pico cell allocation and that of the distance based allocation can be arbitrarily large.

Proof Consider the topology shown in Figure 3.3. There are two macro controllers with available capacity $C_i = p$ and a set of $2p$ pico cells with unit requirements of C-plane traffic is shown in pairs where the first set has handover rate equal to h and the second set has as zero. In the optimal allocation (case(a)) the inter-macro controller traffic is zero since there are no handover between pico cells that are allocated in different macro controllers. In the distance based allocation (case (b)) all pico cells will be connected to the closest macro controller, i.e. the one with distance d1. In that case, the inter-macro controller handover cost would be $\frac{p}{2} \cdot h$ which can be arbitrary large with the available capacity C_i. □

In order to capture the benefits of the proposed set of solutions compared to the baseline distance based small cell association is insightful to evaluate the schemes under different scenarios with variable congestion episodes levels. To this end, we define the congestion level in the network using the following variable Ω, which expresses the ratio between the total C-plane requirements of the small cells in a given geographical area (which depends on the number of small cells as well as the number of mobile users) and the aggregate capacity of the macro controllers.

$$\Omega = \frac{\sum_{i \in \mathbf{P}} g_i}{\sum_{k \in \mathbf{M}} C_k} \leq 1 \tag{3.16}$$

The Eq (3.16) reveals the congestion level of the problem. If the value of Ω is approaching to 1 then it reveals that network congestion level becomes very high. In terms of QoS, a very high congestion level in control signal will cause delay in signalling between Control plane and data plane. This would cause issues in

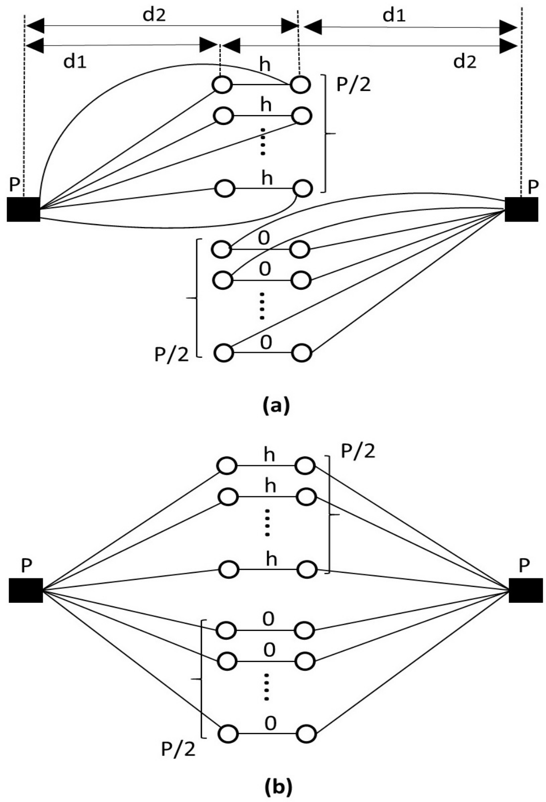


Fig. 3.3 The scenario depicts $2p$ pico cells that have one unit rate for C-plane requirement and the handover rate between half of the pico cell pairs is h and zero for the rest. The two macro controllers have available capacity $C_k = p$. Case (a) shows optimal allocation with zero inter-macro controller handover overhead, case (b) worst-case allocation of pico cells to macro controllers.

setting up connection between pico cells and its serving users, which potentially reduces the QoS of users. Such problem can be even worse when there are more users trying to connect to networks, which increases the burden in data plane, and accordingly in the control plane.

3.5 Numerical Investigation

In this section, a wide set of numerical investigations to test the performance of the proposed optimization under various network scenarios and network congestion levels is provided. The system model parametrization and assumptions are built mainly on the proposed C/U split phantom cell concept work that has been presented in [119]. The topology where applying both optimizations and baseline distance based small cell association is assumed as a 3 overlapped MBSs network and the radius of each MBS hexagon area is 500m. The control plane of macro cell is equipped with the control plane interface based on the Evolved Packet Core (EPC) entities. The number of SBSs in the investigations ranged in the following discrete values 10, 15 and 20. These SBSs are placed randomly in the edge cell overlapped area between different MBSs. In the distance based SBS association, the allocation of SBSs is established based on the minimum distance of that SBS with the different MBSs. Therefore, SBSs in that case are associated to the nearest MBS without considering handover rate and/or MBS congestion levels.

In the numerical investigations, MBSs have available capacity C_k and SBSs have the requirement on control data rate g_i as mentioned in Table 3.1 (we note that all range of values are uniformly distributed). As estimated from mobility performance, the per edge user data rate can reach approximately 0.9 Mbps. Therefore, we assume the range of per user control signaling requirements is uniformly distributed between 0.6-0.9 Mbps. Based on these values and average number of users per small cell we assume that each small cell has a range of control signaling requirements, g_i, distributed between 0.6-4.5 Mbps uniformly in our simulations. The Table 3.1 summarizes the main system level parameters that have been used.

3.5.1 Control Plane Handover Signal Optimization

The inter macro cell handover rates comparison between different scenarios is shown in Fig.3.4 for the proposed optimization problem and the distance based

Table 3.1 SIMULATION PARAMETERS

Parameters	**Values**
Number of macro cells (M)	3
Number of small cells (P)	10,15,20
Macro cell radius	500 m
small cell radius	50 m
C-plane requirements per UE	0.6 - 0.9 Mbps
Number of UE per small cell	1 - 5
C-plane requirement per small cell (g_i)	0.6 - 4.5 Mbps
Macro cell available capacity for control (C_k)	10 - 35 Mbps
Handover rate (h_{ij})	0 - 1 Kbps
Congestion level (Ω)	0 - 1

small cell association heuristic. More specifically, a network scenario with different number of small cells and 3 macro cell controllers is considered and the figure depicts the total handover rates between potential handover small cells in 3 macro cells coverages. The proposed optimization problem provides an average of 37% improvement compared to the distance based heuristic across all network scenarios. This result indicates that the proposed optimization framework provides a significant decrease of the inter macro cell handover rates which translates to signalling overhead and latency on the handover completion.

3.5.2 Control Signal Load Balancing Optimization

The performance evaluation of the proposed load balancing optimization strategy is shown in Figure below. The subfigure 3.5(a) indicates the utilization level on the macro cell which has the minimal control traffic for the proposed load balancing scheme and the distance based heuristic for different network scenarios. In addition, subfigure 3.5(b) indicates the unbalance level in the network for different scenarios. As can be observed, the proposed technique provides a performance im-

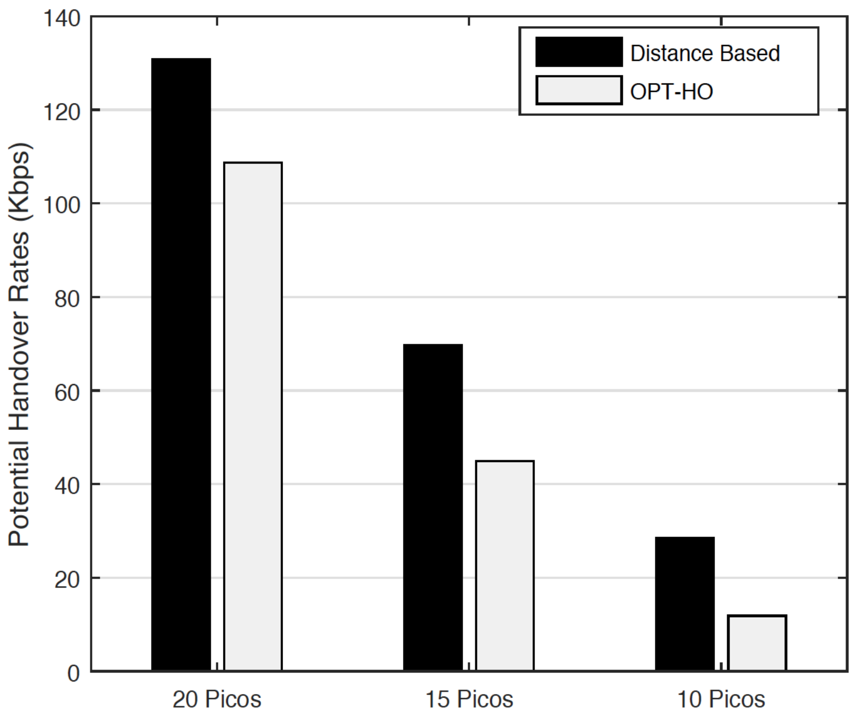
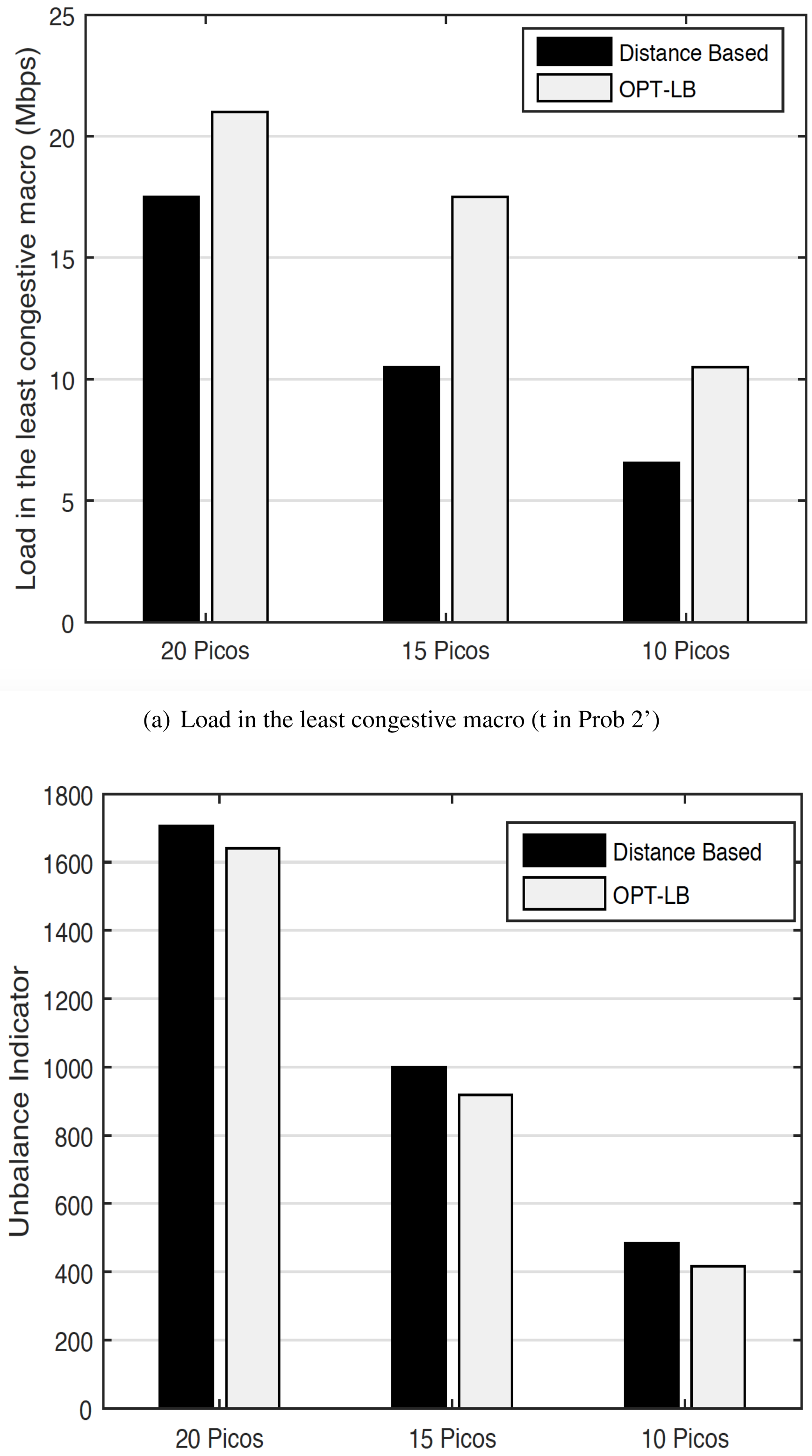


Fig. 3.4 Performance on inter macro cells handover rates

provement across all network scenarios in 3.5(a) and 3.5(b) approximately as 49% and 10% respectively. From these results, it is evident that the proposed technique that strive to distribute load across macro controllers manages successfully to keep the minimal load consumption larger than that of the distance based scheme and always achieve a more balanced distribution. In Fig.3.6 we show the control plane load for each macro cell focusing on the 15 small cells network scenario. The control plane rate requirement for each small cell assume in this evaluation is 3.5 Mbps and the available C-plane capacity for each macro is assumed to be 30 Mbps. As shown in the figure, the proposed optimization algorithm manages to successfully distribute control plane small cell association equally to different macro cells (5 small cells for each macro) compared to the performance achieved by the



(a) Load in the least congestive macro (t in Prob 2')

(b) Unbalance level for whole network

Fig. 3.5 Performance on inter macro cells load balancing

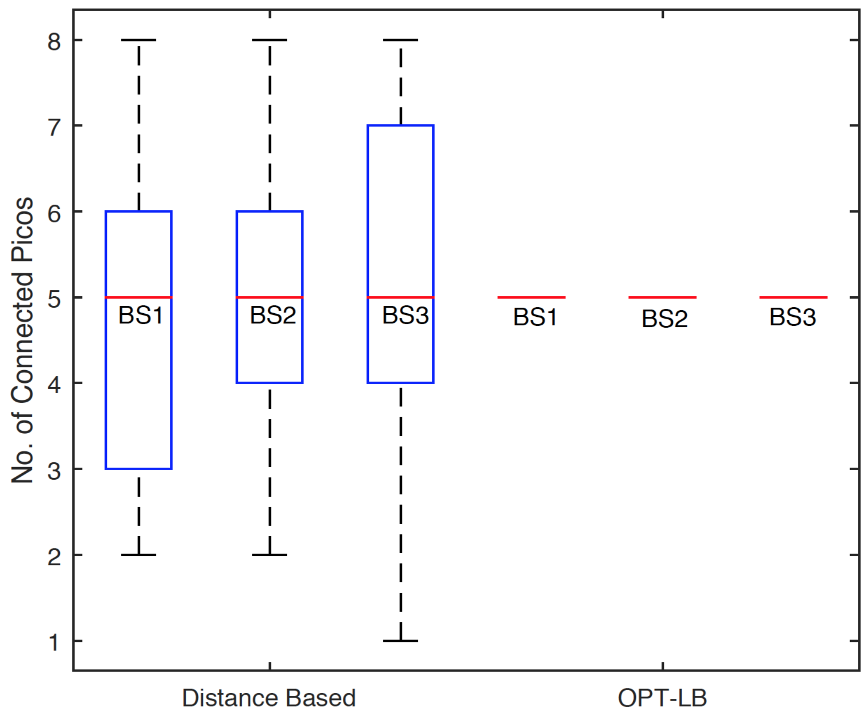


Fig. 3.6 Picos distribution for 3 macros controlling 15 picos

distance based association which entails a significantly more unbalanced C-plane load distribution (small cells associations for each macro varies unpredictably).

3.5.3 Performance Summary and Discussion on Joint Optimization

Finally, Table 3.2 provides a summary of the numerical investigations and shows the enhancement in handover optimization and load balancing respectively compared to distance based solution. As can be seen, with the increasing of the congestion level, both of optimization algorithms would perform weaker than which in less congestive network.

Furthermore, it is necessary to discuss the proposed weighted optimization problem that allows for flexible network operation by suitably choosing the weights

Table 3.2 OPTIMIZATION ENHANCEMENT

Enhancement	20 picos	15 picos	10 picos
Handover	16.80%	35.54%	58.42%
Load Balancing	3.28%	10.03%	15.19%

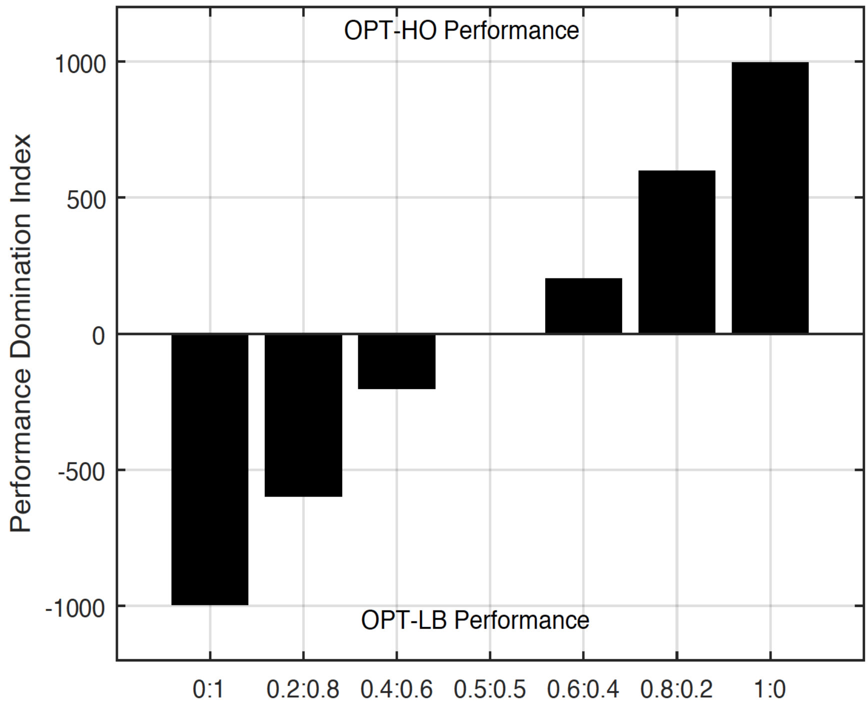


Fig. 3.7 Optimization Leverage (ω_1:ω_2)

ω_i for two parts of optimization. In a nutshell, the proposed optimization (Prob 3) aims to allow flexibility to enforce various levels of load balancing and inter macro cell handover rate reduction. As shown in Fig.3.7, because of the normalization performed on objective function, with the varying weights ω_1 and ω_2 different levels of contribution from handover reduction and load balancing are shown as expected which is in a form of standard Pareto frontier.

3.6 Further Discussion on the Proposed Scheme

3.6.1 Advantages

The key advantage of the proposed approach is that it allows to consistently provide optimal solutions to both overhead reduction problem and load congestion (balancing) in the control plane traffic. In that respect, provides an upper bound indicator on the performance improvement that can be achieved in the network. We refer to the numerical investigations section of the paper that detail the main benefits of the proposed scheme in terms of performance improvement that can be achieved (load balancing, with a small penalty on the aggregated throughput).

3.6.2 Disadvantages, Further Issues to be Considered

However, there are still some disadvantages need to be raised up. First of all is the practical consideration of testing the algorithms. The experimental results are based on extended Monte Carlo simulations using MATLAB but care should be taken to translate those with respect to more realistic network cases in terms for example small cell deployments and actual users handover that depend on their mobility pattern. Clearly those can vary significantly since they depend very much on the location and user distribution.

The proposed optimization per se, requires *average handover rates* between adjacent cells to be monitored and being readily available to the algorithm and *cell load factors*. This type of information will need to be communicated from the small cells to the network controller (and being periodically updated). The evaluations was based on assumptions regarding the values of those metrics but to get a more realistic set of results, the algorithms will need to be tested on real world traces. For example, average user handover rates between adjacent cells could be acquired from network operators but it is in general difficult to have access to such data especially for small cells deployments of high density.

Furthermore, the proposed mathematical programming setting is in general not a scalable framework due to the inherent non-polynomial complexity of the problem at hand. Numerical investigations reveal that it is possible to run such a framework for low to medium network size instances but might lead to increased running times for large network instances. Therefore, heuristics and/or greedy solution methodologies need to be defined in order to allow scale free operation for any network instance. One issue that also need to be considered is if there are any conflicts between the proposed algorithms and other network functions in network entities (such as for example admission control, even though this functionality is not being investigated within 5G NORMA). These issues will be more clearly shown and/or clarified via a signalling procedure that will detail how this framework can be considered as a building block of a generic SDM (Software Defined Mobile) controller in the network.

3.7 Conclusions

Over the next couple of years we expect a high degree of cell densification in cellular networks that will result in an explosion on the utilization of higher spectrum frequencies (mmWave) in order to increase overall spatial capacity. In that setting, efficient, low overhead and complexity network orchestration is emerging as an important tenet of emerging wireless networks. The logical/physical decoupling/split of control and user planes (C/U) is envisioned as a key architectural element to provide the required flexibility in managing small cell networks. To this end, in this paper we detail network optimization algorithms for a generic C/U split architecture with the emphasis being on control plane load balancing and C-plane load reduction by taking explicitly into account the inter macro controller handover of mobile users as well as congestion levels in different macro cells. In C/U split architectures we therefore aim to ameliorate network congestion episodes where the performance of the network might potentially be limited by the macro-cell control plane capacity rather than the high capacity small cells. Therefore, optimization algorithms for easing the congestion level of the control plane at the macro cells is an important element to ensure a high performance in emerging wireless networks.

Chapter 4

INTER-TENANT RESOURCE SHARING AND POWER ALLOCATION IN VIRTUAL NETWORK

Nowadays the network virtualization is brought to the spotlight due to the booming user requirements in wireless communication network. However, the potential price for facilitating network slicing in a multi-tenant virtual network is the underutilization of the treasured wireless network resources because of the different tenant requirements and dynamic traffic episode. A potential way to avoid such sacrifice of radio resources is to allow further inter-tenant sharing between tenants. To this end, this work proposes an optimization framework by bringing a flexible inter-tenant resource sharing scheme embedded with transmission power control to aggressively improve network capacity, user data rate and energy efficiency. More specifically we define two resource sharing mechanisms called Tight Coupling

(TX) and Loose Coupling (LX) and formulate them in a form of Mixed Integer Linear Programming (MILP) problem. Furthermore, two Resource and Power Joint Allocation (RPJA) algorithms are designed to solve the optimization problem in polynomial time. Under the 3GPP standards and certain assumptions, numerical investigations have been conducted demonstrating a significant gain in network throughput, individual user rate and energy efficiency, comparing with classic network slicing methods and constant power resource sharing algorithms.

4.1 Introduction

Over the past few years, a significant attention has been paid to defining and architecturing 5G communication networks, which aims to transfer the ordinary cellular networks to the denser deployed heterogeneous networks (HetNets). These emerging networks are expected to provide highly increased data rates but more importantly to efficiently and cost-effectively support new services and vertical market integration to create in that sense an ecosystem for both technical as well as business innovation [5]. However, the densely deployed multi-tier networks that combining both MBS and SBS can cause significant operating cost in infrastructure and chaos in network management/cooperation. On this point, enabling the logically isolated, coexisted and shareable virtual network within the physical infrastructure is seemed as a compulsory component to future 5G network. In terms of wireless resources, resource slicing and potential sharing among multi-SPs/tenants is the key to support diversity and capacity in network. By combining these concepts and technologies, the costs reduction of InPs (both CAPEX and OPEX), and simpler cooperation between SPs can be achieved.

Hand in hand with the aforementioned techniques and benefits, a set of challenges also arrive when bringing virtualization functionalities to a dense multi-tier RAN. Although the Third-Generation Partnership Project (3GPP)[119] has standardized certain functionalities to motivate multiple SP to share a cellular network, there are still limitations to defining the specific resource assignment among virtual slices and detailed implementation of functionalities. Summarized in Chapter 2, the research gaps are still opened for the following areas: (1) how could radio resources be shared and reused intelligently between tenants based on customized tenancy agreement, (2) how could radio resources be allocated dynamically and flexibly in virtual network with varying traffic demand (especially in overload scenarios) and (3) how could extra controlling signal that occurs due to the communications between SP and InP be handled to ensure the normal operation of network. In the view of wireless architecture, these issues become more challenging when applying multi-tier HetNets in which overlapped and densely deployed base stations are suffering from more sources of interference.

To this end, by studying the present state of the art and principle of network architecturing, we hereby propose an optimization framework to flexibly share and reuse the OFDMA (Orthogonal Frequency Division Multiple Access) based PRB among tenants in a virtualized wireless network. Furthermore, to aggressively improve the overall network capacity and avoid overwhelming interference, we combine the intelligent resource reuse approach with transmission power controlling technique and shed the light on different degrees of resource sharing. In technical aspect, this work defines a resource reuse maximization problem in a form of MILP, and provides two low complexity algorithms to achieve improvement in system throughput, per user rate and energy efficiency. To serve the principle of network architecturing, SDN and C-RAN are introduced in order to alleviate the chaos of controlling signals and inter-tier interference.

Based on the current state of the art and related works, from our best knowledge there is no work considering that enables further resource reuse of orthogonal resources by introducing a scalable degrees of inter-tenant sharing with endurable interference. The main contributions of this chapter include (1) proposal of an optimization framework that offers inter-tenant resource reuse/allocation with different sharing flexibility and operation complexity (LX and TX) in a virtual wireless network environment, (2) an illustration of practical implementation of proposal in SDN network enabled by specified SDM [120], (3) an investigation on binding variable transmission power to motivate further inter-tenant resource reuse with benefit of improved energy efficiency, and (4) two heuristic based algorithms (RPJA-h and RPJA-adv) that can be embedded in network controllers to provide efficient sub-optimal solutions to the optimization problem.

The rest of this chapter is organized as following: Section 4.2 presents detailed explanation of SDM, where our scheme and related algorithms can be fully embedded, and two scalable sharing operation mechanisms, the LX and TX. In Section 4.3, the realistic network topology used by the framework is introduced and optimization problem is formulated accordingly. Section 4.4 introduces the proposed algorithms that eliminate the complexity and dimension. Section 4.5 illustrates the details of the numerical investigations by comparing proposed scheme with classic methods and, at the end Section 4.6 concludes the work.

4.2 Inter-Tenant Sharing and Power Allocation Framework (LX & TX)

This section illustrates the different operations of inter-tenant sharing and power allocation proposed in this chapter. To specify the operations in the horizon of network engineering, the software defined mobile (SDM) controllers are also explained.

4.2.1 Software Defined Mobile (SDM) Controller

To satisfy the need of performing scalable inter-tenant resource sharing in a network architecturing horizon, the specified SDM controllers are introduced as the key elements to virtual networks. The SDM controller is usually defined in four aspects: topology, resource, function and deployment [120]. In resource aspect, the main function SDM provides is to slice the bandwidth of virtual network (also called SDM orchestrator in this case) and realize real-time scheduling and resource management for different network slices. The slice, as introduced in section 2.3, is deemed as a piece of network resources which is available to be assigned to a SP and its users. As introduced in To motivate high flexibility in resource management, various organizations have proposed different ideas of 5G defined virtual controllers that can be embedded into future networks. Broadly speaking, ETSI NFV provides a very precise architectural framework for managing and orchestrating virtualized resources that relate to network functions as well services of an operator. SDN, on the other hand, can be deemed as a more generic framework compared to NFV and in that respect an ETSI NFV architecture can utilize the services of SDN to provide a programmable platform for establishing links between the various VNFs in the sense of routing and overall policy based management. At the moment, ETSI NFV is in the process of specifying the interfaces for allowing interworking between SDN controllers with the NFV MANO system, and various options have already been previously defined in ETSI standard [121].

Since 5G NORMA has been closely working with Centre for Telecommunications research in King's College London, their proposed pair of distinguished SDM controllers: SDM-C and SDM-X, is introduced in this section as example to illustrate how the optimization framework in this chapter implement. However, the proposed set of techniques are generic enough to allow implementation in other potential architectures that support resource sharing in emerging 5G networks defined in ETSI standard.

In definition, SDM-C (Software Defined Mobile Controller) refers to the generic controller operating in dedicated control layer that provides per slice based management, which is also called intra-tenant/slice controller. SDM-C normally focuses on VNFs management and radio resources association within individual network slice. With the given amount of assigned resources, SDM-C can perform instant reconfiguration and optimization on these resources at any point of the time scale.

On the other hand, SDM-X (Software Defined Mobile Coordinator) refers to the inter-slice controller operating in common control layer which is in charge of inter-slice resource management. The main work of SDM-X is to implement the agreed network sharing policies between slices/tenants. In general, it tightly interacts with all SDM-C in control layer to synchronize the information related to VNFs and resource usage all the time. Due to the isolation nature of virtual slices, the existence of SDM-X is the key to guarantee the overall network requirement among all tenants and avoid unacceptable conflicts especially when network slices are agreed to share their assigned resources.

During the process of inter-tenant resource sharing, both of the controllers can work either separately or congruently depending on the service request or estimated network interference. Regarding to this, the inter-tenant resource sharing framework shall operate under a proper cooperation between SDM-X and SDM-C. With the programmability and softwarization of SDN, optimization and well-designed algorithms like the proposal in this chapter could be embedded into the virtualized control layers to improve overall network performance and avoid potential inter-tenant conflicts in resource assignment.

4.2.2 Operation Mechanisms: Loose Coupling (LX) and Tight Coupling (TX)

Based on the virtual controllers described previously, this part hereby details two novel operation mechanisms under the framework of proposed inter-tenant resource sharing and variable power allocation: Loose Coupling (LX) and Tight Coupling (TX).

Loose Coupling (LX)

LX operation refers to the scenario where per tenant resource partition and inter-tenant shareable resources are pre-defined at the beginning of system operation. Therefore inter-slice control functionalities provided by SDM-X need to be set up before the operation of optimization or algorithmic framework. Once the operation begins, LX assigns PRBs to users and allocates a proper transmission power within a defined power range to the assigned PRBs based on optimal reuse principle.

By choosing a proper transmission power, more PRBs reuse can be engaged and interference can be further limited.

Tight Coupling (TX)

This case is recognized when resources of individual tenant and those shared ones are decided dynamically during the operation. Computationally TX is the most complex one in operation and it requires the fully functional overlapped between SDM-C and SDM-X. Compared to LX, TX only has the information about the amount of shareable resources before operation, and which PRBs shared in each tenant can be consistently changed according to resource re-usability, system interference and traffic conditions. The variable transmission power is also available in TX to engage the reuse of resources.

Toy Example Illustration

In this example, assume there are only 6 PRBs can be utilized in BBU pool and two tenants exist to share the whole network. The figure shows different part of control layer: the common control layer and dedicated control layer. In terms of resource sharing, in LX operation, the SDM-C controller assigns the divided bandwidth to each tenant meanwhile the SDM-X controller optimizes the resource allocation on each slice and the reuse of pre-defined inter-tenant sharable resources. On the other hand, in TX operation, the SDM-X controller takes all resources into account to find optimal subset for sharing through operating time scale. This distinction is depicted in Fig.4.1 that the chosen PRBs for inter-tenant sharing is unknown without acknowledgement of traffic scenario. In essence, TX operation is considered as the most general and complicated case of LX by including all available resources in the shared set only with resource amount limitation.

Therefore, in the eye of network engineering, the trade off or implementation preference between TX and LX can be evaluated based on both sharing flexibility, interference control ability and system operation complexity.

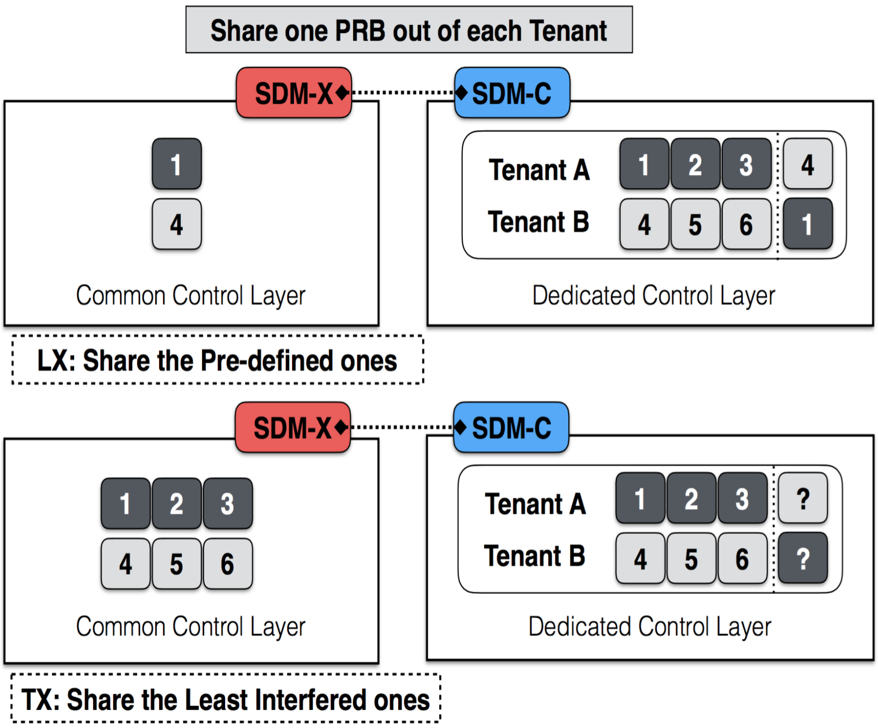


Fig. 4.1 Two sharing mechanisms explanation: in this example each tenant shares only one of its PRBs. In LX operation, two sharing PRBs (1 & 4) are specified manually but, in TX operation, the sharing PRBs are decided based on the re-usability and interference dynamically during operation

4.3 System Model and Problem Formulation

This work considers a C/U split SDN architecture where SBSs are densely deployed in a centralized MBS controlling area. For this reason, we can ignore the inter-tier interference to simplify the model. In C/U split network, the data planes (the SBSs) will be only focused on forwarding user data rate while control planes (the MBSs) provide control signal links. All SBSs in network transmit over a common frequency band (frequency reuse is allowed) but which is differentiated from the one used by MBS. Geographically we assume a number of K SBSs distributed in the MBS controlling area based on a Poisson Point Process (PPP) recognized as user hot zones, and a number of N UEs are randomly located in the hot zone SBSs. To serve the arrival traffic, control plane creates a centralized resource pool based on the bandwidth of data forwarding signal and SPs/tenants registered in this network can require virtual resources from the pool. In this case, the SPs enable to serve their subscribers/ end users via the aggressive data forwarding SBSs with the best signal link. The number of PRBs M in the pool is directly depended on system bandwidth of data forwarding and only downlink transmission is considered in this framework.

An example of network topology is shown in Fig.4.2. The available PRBs and UEs of two tenants (A and B) are colored blue and red, respectively. In the centralized BBU resource pool, two marked PRBs are those chosen to be shared between tenants in this example without specifying LX or TX operation.

4.3.1 Preliminaries

To model the resource association and power allocation problem, the optimization framework defines the following binary decision variables as preliminaries.

$$x_{rt} = \begin{cases} 1 & \text{if PRB } r \text{ is assigned to tenant } t. \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

$$y_{irtl} = \begin{cases} 1 & \text{if user} i \text{ of tenant} t \text{ uses PRB} r \text{ with power level} l. \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

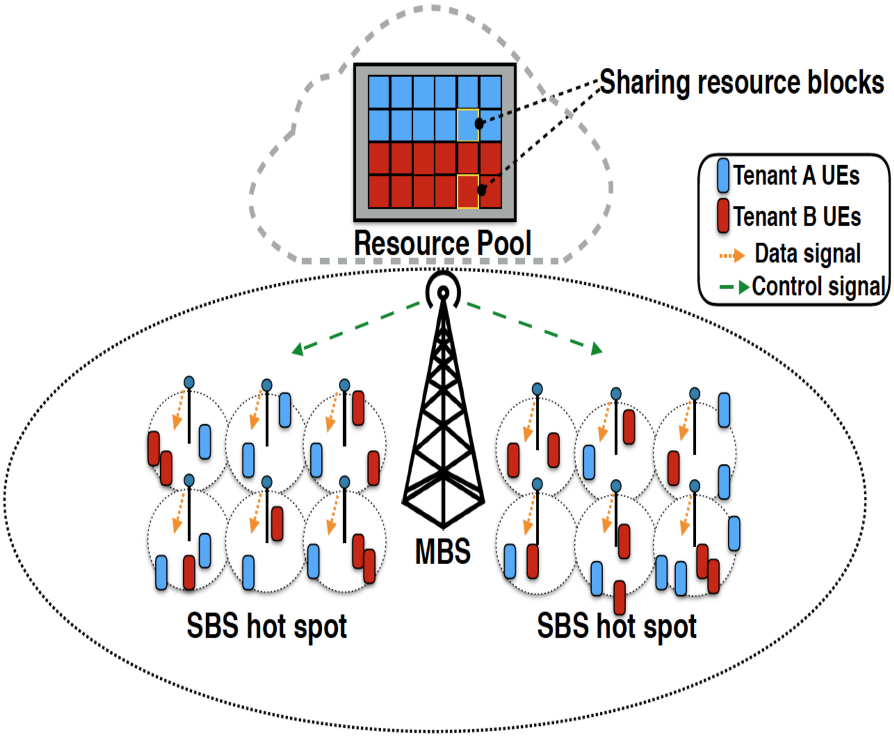


Fig. 4.2 An example of C/U split RAN architecture with traffic from two tenants: MBS controls SBSs which forward user data in hot zones and builds the centralized resource pool

Notations **I**, **C**, **T** and **P** are used to indicate the set of UEs, PRBs, tenants and power level respectively. In this paper, only mathematical set and variables are written in bold style. Besides, this framework denotes discrete power level by l [122][123] to embrace the nature of integer linear program, linking to a power value $p^{i'}_{irtl}$, which represents a controllable transmission power provided by the base station serving user i. Research in [122] shows the details of how a PRB can be allocated variable power by using discrete power level. As mentioned before, so far the system model has $|\mathbf{C}| = M$ PRBs in resource pool, $|\mathbf{P}|$ power levels and $|\mathbf{B}| = K$ SBSs (in a set $\mathbf{B}$) waiting to serve a number of $|\mathbf{I}| = N$ users from number of $|\mathbf{T}|$ tenants, then the potential maximum power of each PRB is $p^{max}_{irtl} = P_{tx}/M$. The notation P_{tx} is denoted as the maximum transmission power of a SBS, in our case all SBSs have the same maximum transmission power. Based on this, the variable power can be allocated to each PRB is defined as,

$$p_{irtl}^{i'} \leq p_{irtl}^{max} \tag{4.3}$$

Furthermore, we define the following indicator to express the orthogonal resource reuse principle.

$$\phi_{ij}^{tt'} = \begin{cases} 1 & \text{if users i,j of tenants t and t' in different BSs.} \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

In this case, if users i and j are from the same tenant then $t = t'$. Regarding to the channel gains on downlink transmissions, the framework denotes by $g_{irtl}^{i'}$ the link gain between the serving base station i' and user i using PRB r. Also, with $g_{irtl}^{j'}$ the framework denotes that link gain between the serving base station of user j (base station j') and user i as the interference link. Here is no cross-tier interference due to the no data forwarding nature of MBS. In that respect, the SINR for a PRB-UE pair can be estimated by following expression,

$$\gamma_{irtl} = \frac{g_{irtl}^{i'} p_{irtl}^{i'} y_{irtl}}{\sum_{j \in \mathbf{I'}} \sum_{t' \in \mathbf{T}} \sum_{l \in \mathbf{P}} \phi_{ij}^{tt'} g_{irtl}^{j'} p_{irtl}^{j'} y_{jrt'l} + I_{noise}} \tag{4.5}$$

The term γ_{irtl} and I_{noise} indicates the SINR and the background noise of the channel respectively. To be noted, the set $\mathbf{I'}$ represents the user set without UE i.

Based on Eq.(4.5), the achievable data rate of UE i by taking PRB r can be approximated by the Shannon Capacity Formula,

$$R_{irtl}(y_{irtl}) = \Delta f log_2(1 + \gamma_{irtl}) \tag{4.6}$$

where Δf is the LTE-based frequency space for a PRB [124].

Comparing to previous research [125][126] using the aggregated energy as the energy measurement metric, this optimization framework prefers to use Energy Efficiency (EE) since it indicates both the energy consumption and user data rate [127][128][129]. Based on the cumulated energy consumption and aggregated user data rate, the EE in this work is defined as,

$$EE = \frac{\sum_{i\in\mathbf{I}}\sum_{r\in\mathbf{C}}\sum_{t\in\mathbf{T}}\sum_{l\in\mathbf{P}} R_{irtl}(y_{irtl})}{\sum_{i\in\mathbf{I}}\sum_{r\in\mathbf{C}}\sum_{t\in\mathbf{T}}\sum_{l\in\mathbf{P}} p^{i'}_{irtl} + \sum_{k\in\mathbf{B}} P^{k}_{SBS}} \tag{4.7}$$

The term P^{k}_{SBS} represents the circuit/operation power of each SBS [128][129]. The unit of EE is bit/Joule or Mbit/Joule.

4.3.2 Problem Formulation

Based on the preliminaries, a resource reuse maximization problem is defined as follows,

$$[\textbf{Problem I}] \ \max \sum_{i\in\mathbf{I}}\sum_{r\in\mathbf{C}}\sum_{t\in\mathbf{T}}\sum_{l\in\mathbf{P}} y_{irtl} \tag{4.8}$$

s.t.

$$\sum_{r\in\mathbf{C}} x_{rt} \geq n_t, \ \forall t \in \mathbf{T} \tag{4.9a}$$

$$\sum_{t\in\mathbf{T}} x_{rt} \leq \beta, \ \forall r \in \mathbf{C} \tag{4.9b}$$

$$\sum_{r\in\mathbf{C}}\sum_{l\in\mathbf{P}} y_{irtl} \leq \delta, \ \forall i \in \mathbf{I}, \forall t \in \mathbf{T} \tag{4.9c}$$

$$\sum_{r\in\mathbf{C}}\sum_{l\in\mathbf{P}} y_{irtl} \geq 1, \ \forall i \in \mathbf{I}, \forall t \in \mathbf{T} \tag{4.9d}$$

$$\sum_{r \in \mathbf{C}} \sum_{t \in \mathbf{T}} x_{rt} \leq \sum_{t} n_t + \alpha \tag{4.9e}$$

$$y_{irtl} \leq x_{rt}, \ \forall i \in \mathbf{I}, \forall r \in \mathbf{C}, \forall t \in \mathbf{T}, \forall l \in \mathbf{P} \tag{4.9f}$$

$$\gamma_{irtl}(y_{irtl}) \geq \gamma_{th}, \forall i \in \mathbf{I}, \forall r \in \mathbf{C}, \forall t \in \mathbf{T}, \forall l \in \mathbf{P} \tag{4.9g}$$

$$y_{irtl} + y_{jrtl} \leq 1 + V \cdot \phi_{ij}^{tt'}, \forall i, j \in \mathbf{I}, \forall r \in \mathbf{C}, \forall t \in \mathbf{T}, \forall l \in \mathbf{P} \tag{4.9h}$$

$$x_{rt} \in \{0, 1\} \tag{4.9i}$$

$$y_{irtl} \in \{0, 1\} \tag{4.9j}$$

Constraint (4.9a) ensures that every tenant will be allocated at least n_t orthogonal PRBs as agreed in their contract. Constraint (4.9b) indicates that a PRB can be reused up to β times across all tenants, which is one way to manage network interference. Furthermore, constraint (4.9c) aims to limit the potential intra-tenant interference by ensuring that up to δ PRBs can be associated to one UE per tenant. Constraint (4.9d) indicates that for each UE it should at least take one PRB as a successful channel connection. Constraint (4.9e) ensures that only up to α PRBs can be shared between tenants, this can be set by the InP according to some pre-defined policies or rules. The binding constraint between the decision variables y_{irtl} and x_{rt} is expressed in (4.9f). Constraint (4.9g) expresses the SINR threshold γ_{th} that need to be satisfied in order to use a PRB. Constraint (4.9h) defines that users from the same tenant and different tenants should be allocated a different PRB if they connect to the same SBS (same cell avoidance) and V in this case is an arbitrary large integer. Constraints (4.9i) and (4.9j) remind the binary of variables. To guarantee constraint (4.9g) is always satisfied when $y_{irtl} = 0$, γ_{irtl} is re-written as following,

$$\gamma_{irtl} = \frac{g_{irtl}^{i'} p_{irtl}^{i'} y_{irtl} + V(1 - y_{irtl})}{\sum_{j \in \mathbf{I}'} \sum_{t' \in \mathbf{T}} \sum_{l \in \mathbf{P}} \phi_{ij}^{tt'} g_{irtl}^{j'} p_{irtl}^{j'} y_{jrt'l} + I_{noise}} \tag{4.10}$$

In the defined problem, the framework aims to optimize the amount of reusable resource blocks with SINR restriction rather than directly optimize the data throughput. Even so, the framework can still have promising gain in throughput cause our proposal aims to ultimately reuse inter slice resources compared to present state of the arts. However, this gain might not be guaranteed all the time while compare to the same optimization with constant transmission power (same formulation without power level definition). To this end, a lemma is contributed to shed the light on the issue.

Lemma 1: Under proposed inter-tenant resource sharing framework, the maximum system rate can be possibly provided by the non-reuse case with constant power rather than the reuse case with power control in certain scenarios.

Proof: In this case, a simplified mathematic model is shown to proof the statement in lemma 1. The Eq.(4.11) shows the typical SINR calculation for a user i, where ω_1 and ω_2 are the power controlling ratio of the signaling cell and interfering cell respectively, ranged from 0 to 1. The parameters g_i and $g_{i'}$ are antenna gains from the signaling cell and interfering cell for user. Still, two cell has the same maximum transmission power p^{max}.

$$\gamma_i = \frac{g_i \omega_1 p^{max}}{g_{i'} \omega_2 p^{max} + I_{noise}} \tag{4.11}$$

Assume an extreme scenario where two users are located in two adjacent cells with that UE 1 is located at the edge of cell 1 ended up with an extreme small antenna gain as g_1=1 and UE 2 is located at the center of cell 2 with an extreme high antenna gain as g_2=N, a large positive number. Meanwhile assume that the antenna gain from interfering cell for UE 1 is $g'_1 = 1$ (maximum interference) and for UE 2 is $g'_2 = \frac{1}{N}$. To further simplify the SINR calculation, assume that maximum transmission power p^{max} for both cells is 1 and thermal noise I_{noise} is $\frac{1}{N}$ which is small enough to be ignored. Following the Eq.(4.11), the SINR for each user in this scenario can be expressed as,

$$\gamma_1 = \frac{g_1 \omega_1 p^{max}}{g_1' \omega_2 p^{max} + I_{noise}} \tag{4.12}$$

$$\gamma_2 = \frac{g_2 \omega_2 p^{max}}{g_2' \omega_1 p^{max} + I_{noise}} \tag{4.13}$$

With all the denoted values, Eq.(4.12) and Eq.(4.13) can be expresses as,

$$\begin{aligned} \gamma_1 &= \frac{\omega_1}{\omega_2 + \frac{1}{N}} \\ &= \frac{\omega_1 N}{\omega_2 N + 1} \end{aligned} \tag{4.14}$$

$$\begin{aligned} \gamma_2 &= \frac{\omega_2 N}{\frac{\omega_1}{N} + \frac{1}{N}} \\ &= \frac{\omega_2 N^2}{\omega_1 + 1} \end{aligned} \tag{4.15}$$

Based on the achieved SINR, the system sum rate can be calculated (aggregated rate of two users) as,

$$\begin{aligned} R &= \Delta f[log_2(1+\gamma_1) + log_2(1+\gamma_2)] \\ &= \Delta f[log_2(1+\frac{\omega_1 N}{\omega_2 N + 1}) + log_2(1+\frac{\omega_2 N^2}{\omega_1 + 1})] \end{aligned} \tag{4.16}$$

To compare with non-reuse case, Eq.(4.16) can be rewritten to express the following scenarios: (1) only UE 1 is served (ω_1=1 and ω_2=0) and (2) only UE 2 is served ((ω_1=0 and ω_2=1)).

The achievable rate for ω_1=1 and ω_2=0 scenario is calculated as,

$$R_{(\omega_1=1\&\omega_2=0)} = \Delta f log_2(1+N) \tag{4.17}$$

The achievable rate for ω_1=0 and ω_2=1 scenario is calculated as,

$$R_{(\omega_1=0\&\omega_2=1)} = \Delta f log_2(1+N^2) \tag{4.18}$$

If start from first scenario (ω_1=1 & ω_2=0), higher rate can be obtained compared to the rate calculated in Eq.(4.17) by increasing ω_2 from 0 to 1 in Eq.(4.16). On other hand, if start from the second scenario (ω_1=1 & ω_2=0),lower rate can be obtained compared to the rate calculated in Eq.(4.18) by increasing ω_1 from 0 to 1. In this case, it is true to say that either higher or lower rate can be achieved by re-using the PRBs between users with power control compared to non-reuse with maximum a constant power/ maximum power under proposed optimization framework. The achievable aggregated rate can be deeply influenced by the value of N. □

4.3.3 Mathematical Differences between LX and TX

The above problem formulation is the one describing TX mathematically since the inter-tenant sharing subset is completely decided by the optimization problem itself. There is no pre-defined subset of resources that decide which PRBs are shareable or belong to a certain tenant. With this understanding, this section illustrates the difference between LX and TX in terms of formulation based on their available PRB set:

• **TX Optimization**: Problem I is the exact mathematical expression for TX operation. To be noted, the set of PRBs that users and tenants can use are all the same, which is $\mathbf{C}$.

• **LX Optimization**: LX operation has the same objective function and constraints as the TX operation has except the different PRB set for each tenant t, defined as $\mathbf{C_t}$. In the PRB subset $\mathbf{C_t}$ of a tenant, the pre-defined shareable PRBs are marked manually before operation starts. Regarding to Problem I, the resource set $\mathbf{C}$ shall be replaced by $\mathbf{C_t}$ in all constraints while considering different t.

4.4 Proposed Algorithms: Resource and Power Joint Allocation (RPJA)

In this section, this work proposes two algorithms with different complexity to solve the optimization problem in polynomial time due to the NP-hardness of optimization framework. Both algorithms aim to allocate PRBs and variable power to UEs jointly therefore name them the Resource and Power Joint Allocation algorithms (RPJA). In addition, to offer fair comparison between the variable power optimization and constant power optimization, this section further illustrates resource allocation only version of these two algorithms (called RA algorithms) to realize constant power resource allocation in our previous optimization problem.

Algorithm 1: RPJA-h

Data: Location coordinates of UEs and SBSs; network parameters n_t, β, δ, α and γ_{th}; Interference matrix $H_{N\times N} = \emptyset$, PRBs set **C** and UEs set **I**.
Result: PRBs Association, Power Allocation and Rate Estimation.
Step 1: PRB Association and Power Allocation
for *r:=1 to* $M \in$ **C do**
  • sort elements (ascending) in **I** based on p^{min}_{irtl};
  for *i:=1 to* $N \in$ **I do**
    repeat
      • obtain location and channel information of UE i;
      • allocate power $(1+S)\ p^{min}_{irtl}$ from nearest SBS;
      if *(*$\phi^{tt'}_{ij} = 1$ *&* $\gamma_{irtl} \geq \gamma_{th}$ *for all associated users)* **then**
        • y_{irtl} =1, PRB r is assigned to UE i;
        • x_{rt} =1, Mark PRB r used by tenant t;
        • update matrix $H_{N\times N}$;
        • sort remaining elements (ascending) $i+1$ to N in **I** based on $H_{N\times N}$;
      else
        • y_{irtl} & x_{rt} = 0, no association made
      end
    until δ, β ***or*** $(n_t+\alpha)$ *is reached*;
  end
  • $H_{N\times N} = \emptyset$, clear the interference matrix for next PRB;
end
Step 2: Compute Sum Rate
• $R_h = \sum R_{irtl} y_{irtl}$

4.4.1 RPJA-h and RA-h: Greedy based Heuristic

The proposed optimization problem is classified as a knapsack problem due to the nature of assigning discrete PRBs and power levels. Therefore, to explore the sub-optimal solution in polynomial time, this work firstly proposes a simplified greedy based heuristic algorithm named RPJA-h. The procedure of conducting RPJA-h is shown in the pseudo code Algorithm 1: in step 1, RPJA-h randomly selects one PRB from set **C** and associates it with a UE requiring minimum power to pass its SINR threshold, in this way lowest interference will be generated at this stage. The association is made if all the constraints in optimization problem are satisfied. In terms of power, a minimum power p_{irtl}^{min} with a slack S (a small portion of p_{irtl}^{min}) is allocated to the PRB-UE pair which ensures the connection would not lost because of the future reuse of this PRB. Meanwhile, the potential interference will be quantified and saved to an interference matrix $H_{N\times N}$ in which each interfering user pair is indicated by h_{ij} elementally. The matrix is in a form as follows for a N users system:

$$H_{N\times N} = \begin{bmatrix} 0 & h_{12} & \dots & h_{1N} \\ h_{21} & 0 & \dots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} & h_{N2} & \dots & 0 \end{bmatrix} \tag{4.19}$$

With the updated $H_{N\times N}$, the remaining unassociated UEs will be sorted again based on the present system interference until no more association can be made. This step ends when all PRBs are considered. In step 2, the sum rate of all associations R_h is calculated. This algorithm can also be called a interference based greedy algorithm because the sorting is based on $H_{N\times N}$. To compare with constant power optimization, RPJA-h has another version without applying variable power allocation (PRBs association only) thus named RA-h. The only difference RA-h has in Algorithm 1 is that RA-h allocates maximum transmit power p_{irtl}^{max} to association pair instead of $(1+S)p_{irtl}^{min}$ through the whole process. As can be seen, RPJA-h engages more resource reuse compared to RA-h due to its stricter interference control.

4.4.2 RPJA-adv and RA-adv: Advanced Iterative Heuristic

To improve the performance of RPJA-h, this work further designs an advanced iterative algorithm named RPJA-adv. The procedure of implementing RPJA-adv is illustrated in Algorithm 2. The key difference between RPJA-adv and RPJA-h is that RPJA-adv randomly changes the sorting result in RPJA-h with certain possibility for each iteration z thus better PRB-UE association can be expected.

In order to make such change with a certain possibility, a set of mathematical transformation is proposed accordingly. A parameter named critical decision value d_i' for each UE i is defined as

$$d_i = 1 - \frac{E_i}{E_{max}} \tag{4.20}$$

$$d_i' = \frac{d_i}{\sum_{i \in \mathbf{I}} d_i} \tag{4.21}$$

The E_i indicates the sorting element of corresponding UE i which can be either p_{irtl}^{min} or aggregated interference in $H_{N\times N}$ for this UE, and E_{max} indicates the maximum element among all candidate users. Once d_i' for all users has been decided, a critical selection range for each UE is derived as D_i. The first sorted user has critical selection range as $D_1 \in [0, d_1']$ and $D_i \in [d_{i-1}', d_i' + d_{i-1}']$ is the one for the rest of users. As shown in Algorithm 2, the step 1 begins with uniformly generating a random number ζ between 0 and 1 for each iteration z, then the critical selection range D_i is found for each UE by Eq.(4.20) and Eq.(4.21) based on p_{irtl}^{min}. In this case the sorting of users is carried out depended on the distance between ζ and each D_i. With this sorting result, PRB-UE association will be made and minimum power with a slack will be allocated as in RPJA-h. It is important to note that the same sorting decision with mathematical transformation

Algorithm 2: RPJA-adv

Data: Location coordinates of UEs and SBSs; network parameters n_t, β, δ, α and γ_{th}; Interference matrix $H_{N\times N} = \emptyset$, PRBs set $\mathbf{C}$ and UEs set $\mathbf{I}$.

Result: PRBs Association, Power Allocation and Rate Estimation.

Step 1: PRB association and Power Allocation

for *z:=1 to Z* **do**

- generate a random number ζ between 0 and 1;
- **for** *r:=1 to M* $\in \mathbf{C}$ **do**
  - conduct the transformation in Eq.(4.20) and Eq.(4.21) based on p_{irtl}^{min} and generate D_i for all UEs;
  - sort elements (ascending) in $\mathbf{I}$ based on the distance between ζ and D_i;
  - **for** *i:=1 to N* $\in \mathbf{I}$ **do**
    - **repeat**
      - obtain location and channel information of UE i;
      - allocate power $(1+S)\ p_{irtl}^{min}$ from nearest SBSs;
      - **if** $(\phi_{ij}^{tt'} = 1\ \&\ \gamma_{irtl} \geq \gamma_{th}$ *for all associated users*$)$ **then**
        - y_{irtl} =1, PRB r is assigned to UE i ;
        - x_{rt} =1, Mark PRB r used by tenant t;
        - update matrix $H_{N\times N}$;
        - conduct the transformation in Eq.(4.20) and Eq.(4.21) based on interference in $H_{N\times N}$ and generate D_i for remaining UEs;
        - sort unassociated elements (ascending) in $\mathbf{I}$ based on the distance between ζ and D_i;
      - **else**
        - y_{irtl} & x_{rt} = 0, no association made
      - **end**
    - **until** δ, β ***or*** $(n_t+\alpha)$ *is reached*;
  - **end**
  - $H_{N\times N} = \emptyset$, clear the interference matrix for next PRB;
- **end**
- $R_{(z)} = \sum R_{irtl} y_{irtl}$;

end

Step 2: Provide Best Performed Sum Rate

- $R_{adv} = \max\{R\}$

is conducted based on interference matrix $H_{N \times N}$ for the remaining users until no more association can be made afterwards. At the end of step 1 the sum rate for each iteration z is estimated and saved to database, and in step 2 result of the best sum rate iteration will be generated as final output of RPJA-adv. Similarly, another version of RPJA-adv with constant power allocation is also designed, as RA-adv.

Similar to Problem Formulation, both algorithms shown in this section is based on TX operation. For LX operation, the subset of available PRBs shall be decided based on the tenancy of users, which is known when network controller is obtaining the location and channel information of users.

4.4.3 Complexity

Moreover, the complexity and scalability of algorithms are also outlined. The optimization problem with all set of constraints are in a form of mixed integer liner programming, which is known as NP-Hard problem. In terms of its scalability, the dimension of optimization problem is $|\mathbf{C}| + |\mathbf{T}| + |\mathbf{C}||\mathbf{T}| + \mathbf{2}|\mathbf{I}||\mathbf{T}| + 2|\mathbf{I}||\mathbf{C}||\mathbf{T}||\mathbf{P}| + |\mathbf{I}|^{\mathbf{2}}|\mathbf{C}||\mathbf{T}||\mathbf{P}|$, which makes optimal solution impossible to be computed in real-time networks operation. As can be observed, if the set of resources, tenancies and base stations increase to a large scale, the scalability of proposed problem can be extremely high. On the other hand, the proposed algorithms RPJA-h and RPJA-adv have complexity $\mathscr{O}(|\mathbf{I}||\mathbf{C}|)$ and $\mathscr{O}(|\mathbf{I}||\mathbf{C}|Z)$, respectively, which is deemed can be solved in polynomial time. In terms of complexity, the algorithms are much more efficient than linear programming optimization itself, however, the performance might be compromised in a certain degrees. This complexity and optimality leverage is evaluated in next section.

4.5 Numerical Investigations

The numerical investigation has been conducted based on a wide set of varied scenarios in our research. We firstly explain the system environment and assumptions that realize a C/U split wireless network. Secondly, a set of large scale realistic network simulation is presented, in which we compare the RPJA and RA embedded LX and TX with classic network slicing methods in terms of different performance perspectives. Thirdly, we further illustrate improvement limit, optimality, complexity and scalability of proposed optimization problem and algorithms by introducing a small scale network simulation.

4.5.1 Simulation Scenario and Parameters

The simulated wireless network environment follows the LTE-Advanced network principles proposed by 3GPP [115]. A widely used 10 MHz channel bandwidth is adopted, which provides in total 50 available PRBs in all scenarios we simulated. To realize the C/U split RAN network, a cluster of SBSs is deployed in a MBS (located at the center) controlling area. The number of SBSs varies from 3 to 15 and the deployment follows a Poisson Point Process (PPP). The selected SBSs are standard pico cells that serve traffic in hot zones. We assume the base station transmit with its maximum power in the case without power control and only downlink signal is considered. The performance of proposed optimization framework is sensible to the number of macro cells, pico cells and UEs since these are related to the total network capacity and traffic loads. However, as expected, the optimization performance is linearly related to the number of macro cells and pico cells, but reversely linearly related to the number of UEs. In this numerical investigation, only one scenario is used to illustrate the optimization framework performance as suggested in [130].

A summary of used parameters is concluded in Table 4.1. The custom made simulation has been designed and implemented in MATLAB based on the extensive Monte Carlo iterations. In each iteration, UEs from two tenants are randomly distributed in the hot zones. The SINR threshold is decided by the average channel conditions of all users in each iteration. In addition, according to [130], the general used maximum transmission power of a pico cell is 0.25 W and the circuit power of each is 14.9 W.

Table 4.1 SIMULATION SPECIFICATION AND PARAMETERS [115]

Parameters	Values / Assumptions
Network layout	1 MBS with 3 - 15 SBSs
Cell radius (m)	MBS:1000 SBS: 200
Carrier frequency (GHz)	2
MBS antenna gain (dBi)	14
SBS antenna gain (dBi)	10
Attena configuration	1 Tx for BS, 1 Rx for UE
Thermal noise (dBm/Hz)	-174
System Bandwidth (MHz)	10
No. of PRBs in the pool	50
SBS Path loss (dB)	140.7+36.7log$_{10}$(D) (D in km)
MBS Path loss (dB)	128.1+36.7log$_{10}$(D) (D in km)
Shadowing standard deviation (dB)	6
Max SBS TX power (dBm)	24
SINR threshold γ_{th} (dB)	25
MBS TX power (dBm)	43
Number of UEs	50 - 150
Max No. of a PRB reuse	5
Max No. of UE requirement for PRBs	2

In order to provide well-rounded results, this work compares the proposed algorithms with two classic virtual resource slicing approaches in virtual network for a two tenants scenario: resource based NVS (NVS-RB) and static reservation (SR). As mentioned in Introduction, the NVS-RB [102] offers certain flexibility in network slicing, it initially reserves a proportion of resources for each slice but automatically transfer resources between slices based on traffic conditions (dynamic resource requirement of each tenant). On the other hand, the SR remains the fixed partition of resources for slices despite the variation of traffic. The advantage of NVS-RB is the self-reacting adjustment for slicing resources based on incoming traffic therefore in all simulation scenarios our LX and TX mechanisms will be embedded on the top of NVS-RB by slicing PRBs in BBU pool before enabling inter-tenant sharing. In conclusion, for all scenarios, the performance evaluation will be compared among the following cases: LX and TX (using RPJA-h, RPJA-adv, RA-h and RA-adv), NVS-RB and SR.

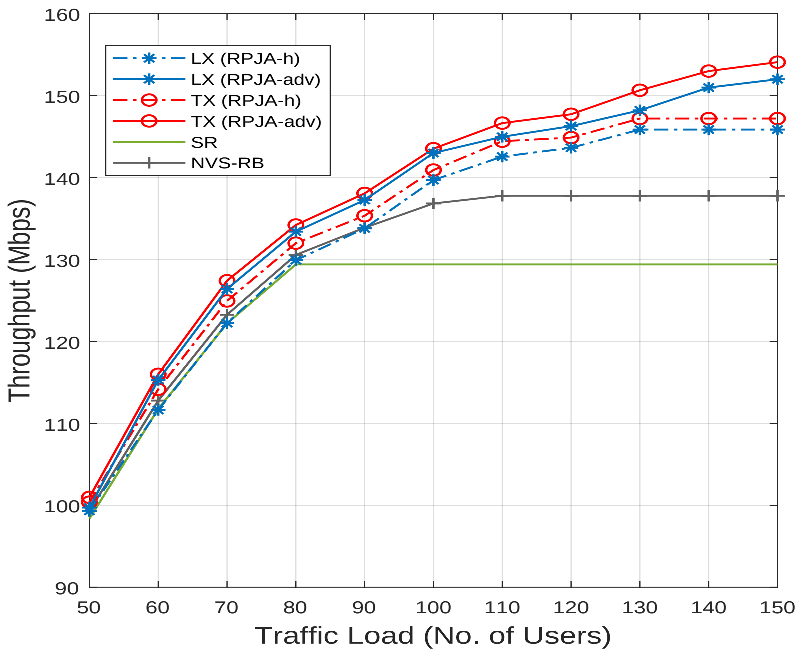
4.5.2 Performance Evaluation

(1) System Throughput Performance

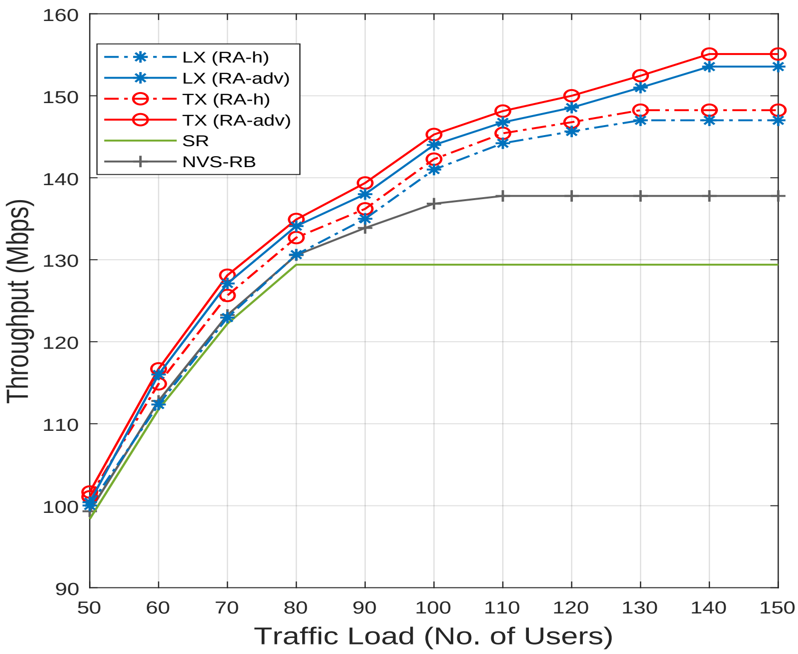
In this large scale simulation, this section shows network performance of 6 SBSs scenario. Besides, the number of UEs varied from 50 to 150 as traffic load. To be noted, we applied 10 PRBs (20% of total resources) as the total amount of inter-tenant sharing PRBs, the α, among two tenants registered in network.

The Fig.4.3(a) and Fig.4.3(b) show the cumulative throughput of variable power allocation (RPJA) and constant power (RA) allocation methods.

As may be observed, the total throughput in all cases increases with the traffic load due to the increasing number of UEs in network. However, the throughput at last comes to a saturated status at certain traffic load since the UEs requirement exceeds the available PRBs in resource pool for all cases. As expected, our proposed LX and TX embedded in both RPJA-adv and RA-adv methods are the last saturated methods in very congested traffic with around 140 UEs. Meanwhile, with slightly loss of maximum throughput of LX and TX embedded in both RPJA-h and RA-h methods is saturated at around 130 UEs scenario, which are still outperformed than SR (saturated at 80) and NVS-RB (saturated at 100).



(a) RPJA throughput



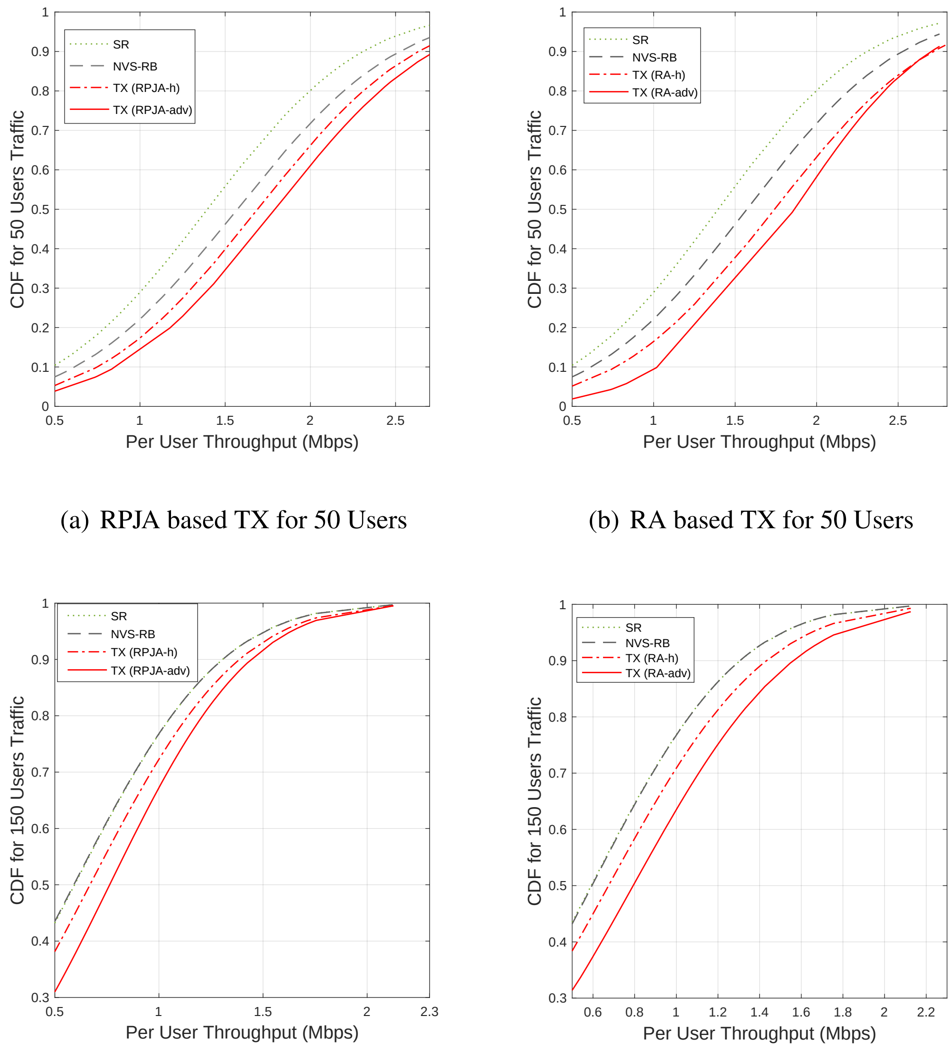
(b) RA throughput

Fig. 4.3 Throughput performance of both RPJA and RA algorithms

In Fig.4.3(a), TX (RPJA-adv), LX (RPJA-adv), TX (RPJA-h) and LX (RPJA-h) achieve the maximum system throughput of 154.1 Mbps, 152 Mbps, 147.2 Mbps and 144.9 Mbps respectively at their saturated point, which translates to a maximum throughput gains by 19.5%, 17.8%, 14.1% and 12.3% compared to SR (129 Mbps), and by 11.6%, 10.07%, 7.4% and 4.9% compared to NVS-RB (138.1 Mbps). With slightly higher gains (see Fig.4.3(b)), the maximum throughput of TX (RA-adv), LX (RA-adv), TX (RA-h) and LX (RA-h) outperformed SR by 20.6%, 18.6%, 15.3% and 14.0%, and NVS-RB by 12.10%, 12.7%, 8.0% and 6.2%. As stated in Lemma 1, the RA embedded methods with constant power can have slightly higher sum rates compared to the RPJA embedded methods with power control due to the dynamic network traffic. In addition, LX usually loses 1-3% throughput compared to TX embedded in all methods due to the flexibility in choosing the shareable PRBs of TX.

(2) Per User Rate Performance

In Fig.4.4(a-d), we present the per user achievable rate based on cumulative distribution function (CDF) analysis. For simplicity of presentation, in this part we only compare TX embedded in all algorithms and the classic methods because the result of LX has the same trend and characteristics as TX. Among these figures, we focus on the CDF percentage of 1.5 Mbps point which can be seemed as a threshold dividing lower rate and higher rate.



(a) RPJA based TX for 50 Users

(b) RA based TX for 50 Users

(c) RPJA based TX for 150 Users

(d) RA based TX for 150 Users

Fig. 4.4 Per user rate performance of TX operation

In terms of this specific point, Fig.4.4(a) and 4.4(b) show that both RPJA and RA based TX operation provide more users with higher data rate ($\geq 1.5Mbps$) compared to classic methods in a low traffic network scenario with 50 users, in which RA-adv, RA-h, RPJA-adv and RPJA-h have 68%, 62%, 65% and 60% respectively of total user population achieving data rate beyond the threshold compared to 44% and 52% of SR and NVS-RB, respectively. This result indicates that our proposed algorithms can not only improve the system throughput but also the individual user rate by engaging aggressive resource reuse between tenants. Similarly, Fig.4.4(c) and Fig.4.4(d) indicate that even in a very congested traffic scenario with 150 users, TX embedded in RA-adv, RA-h, RPJA-adv and RPJA-h also generate 12%, 8%, 8.5% and 7.5% high data rate UEs compared to SR and NVS-RB both with merely 5% high data rate UEs. However, in this extreme congested scenario, not all UEs can be served with a satisfying rate or even connected to the SBSs due to the high interference environment and the lack of resources.

(3) Energy Efficiency Performance

In this part, we evaluate the energy efficiency between different methods as a trade-off for comparing data rate. In Fig.4.5, the EE of all methods with TX operation is illustrated. The reason for using only TX is still that there is no obvious difference in EE results between LX and TX operations in our simulation. We use the box plots to illustrate the variations of EE for all methods through the low traffic load scenario to the high traffic load scenario. Each box itself represents the EE value variation range, and the red band in the box indicates the median in value. Some extreme values are indicated by the red crosses.

Without a doubt, RPJA-adv has the best EE distribution (the box) and RPJA-h, RA-adv, RA-h ranks accordingly behind it. Such results indicate that the RPJA algorithms not only focus on motivating the reuse of PRBs but also lowering transmit power which can still satisfy the signal link. In details, RPJA-adv and PRJA-h provides the EE distribution between 1.32 to 1.53 and 1.29 to 1.50, respectively, compared to the one of RA-adv and RA-h between 1.27 to 1.49 and

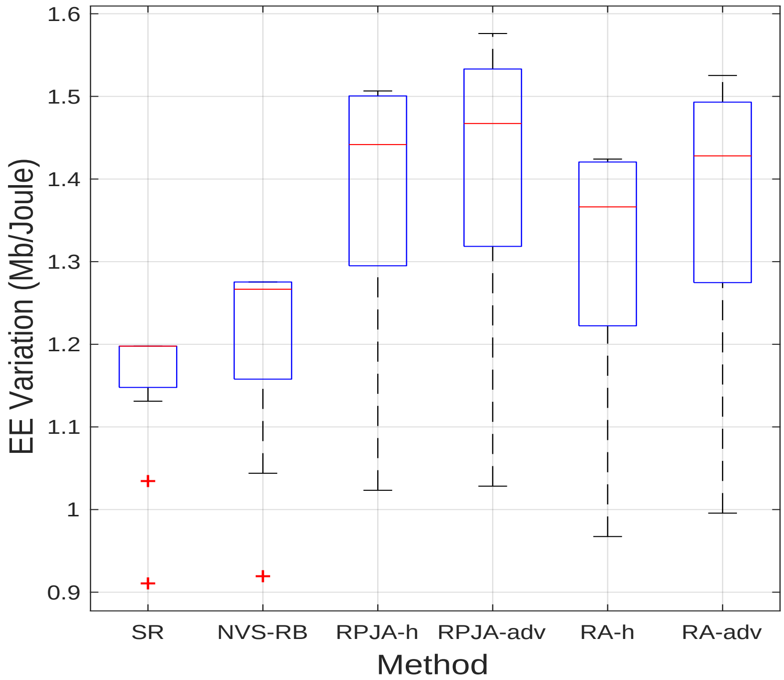
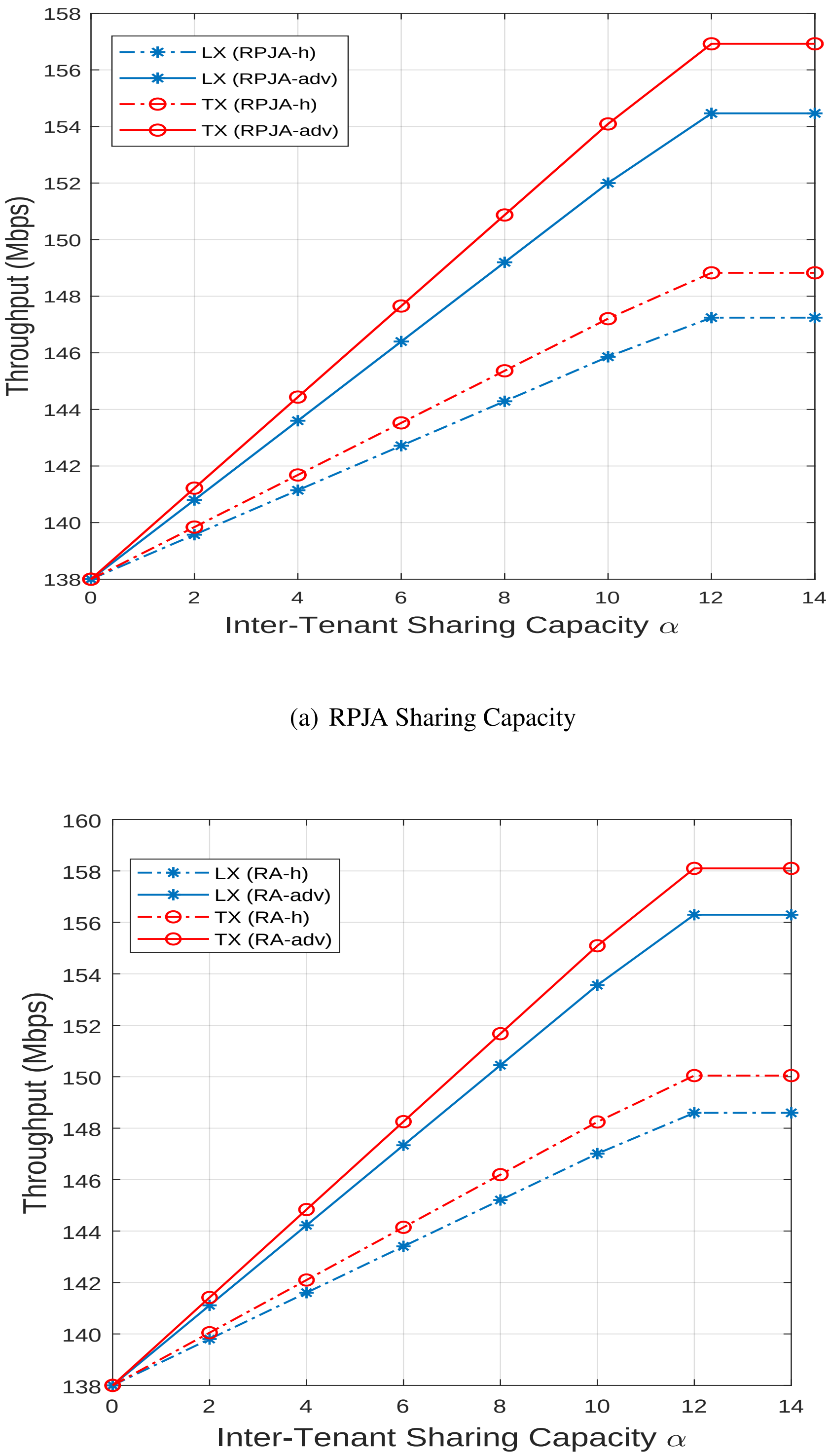


Fig. 4.5 Energy efficiency(EE) advantage of RPJA based TX

1.23 to 1.42, respectively. As may be observed, RA-adv is the one has the closest distribution of RPJA algorithms, which is brought by its advantage in providing high rate. In terms of classic methods we compared with, SR has the worst EE performance with distribution between 1.15 to 1.2, on the other hand, NVS has relatively higher distribution between 1.65 to 1.27 due to the relative higher PRBs reuse efficiency.

(4) Parameters Margin Study

In this set of numerical investigations, we study two parameters that potentially influence the performance of our proposed algorithms. Firstly, Fig.4.6(a) and Fig.4.6(b) show that the change of inter-tenant sharing capacity α, ranged from no sharing to 14 PRBs sharing, can translate to a variation in system throughput. In highly congested network scenario with 150 users, Fig.4.6(a) indicates that the system throughput of RPJA algorithm, TX and LX based RPJA-adv and RPJA-h both increase the network throughput with the increasing sharable number of PRBs. However, they all comes to a saturated status while sharing capacity increased to 12. This happens because the total interference in network reaches its cap therefore no more reuse could happen even though more sharable resources become available. Compared to no inter-tenant sharing scenario ($\alpha = 0$), which has system throughput of 138.1 Mbps, TX(RPJA-adv), LX (RPJA-adv), TX (RPJA-h) and LX (RPJA-h) improve system throughput by 13.7%, 11.7%, 7.8% and 6.6% respectively at their saturated point. The same changing pattern also appear in RA algorithms in result, Compared to no sharing scenario, the maximum gains in throughput provided by TX (RA-adv), LX (RA-adv), TX (RA-h) and LX (RA-h) are 14.5%, 13.1%, 8.7% and 7.4% respectively.



(a) RPJA Sharing Capacity

(b) RA Sharing Capacity

Fig. 4.6 Throughput improvement brought by changing sharing capacity of RPJA and RA for 150 users traffic load

Meanwhile, this work studies the efficiency of proposed RPJA-adv and RA-adv algorithms to locate a near-optimal solution compared to RPJA-h and RA-h. In Fig.4.7, the curve was plotted based on the average number of both RPJA-adv and RA-adv near-optimal solutions compared to greedy heuristic results for different designated iterations Z. It can be easily understood that better solutions are obtained increasingly with the increasing iteration times from initially 50 iterations until 400 iterations is reached. The maximum number of founded near-optimal solutions is around 21 in average and it is unchanged since 400 iterations mainly because of the searching limitation of RPJA-adv and RA-adv.

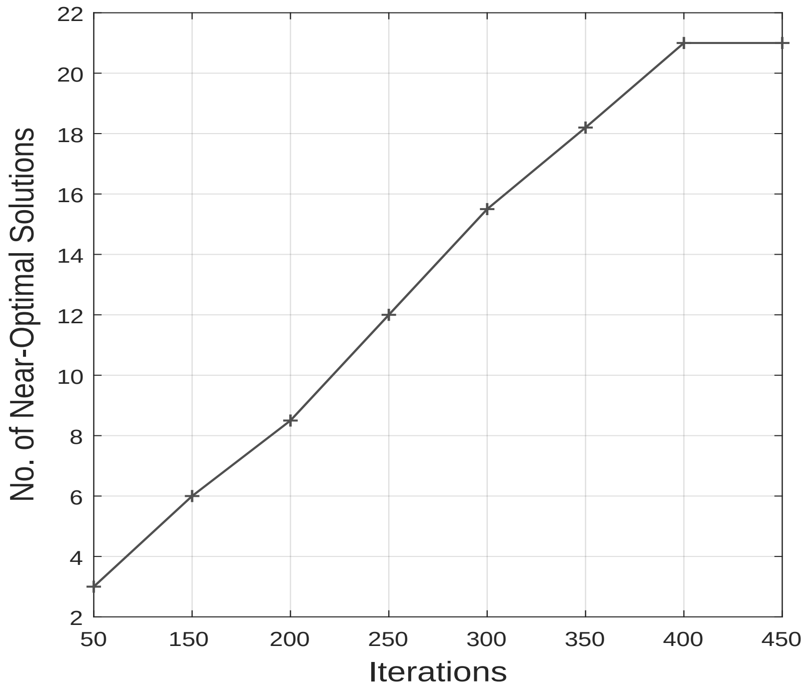
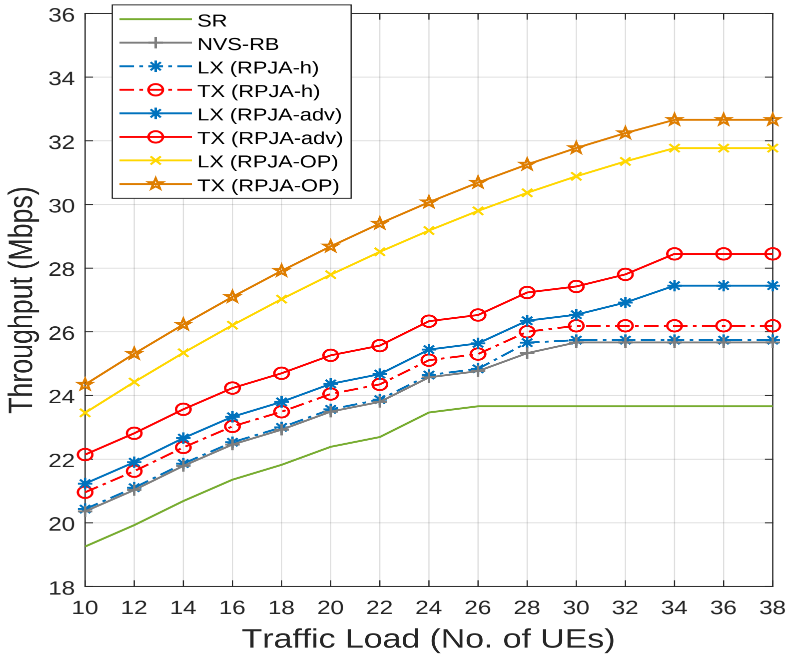


Fig. 4.7 Iterations applied by RPJA-adv and RA-adv algorithms for Searching the Solutions

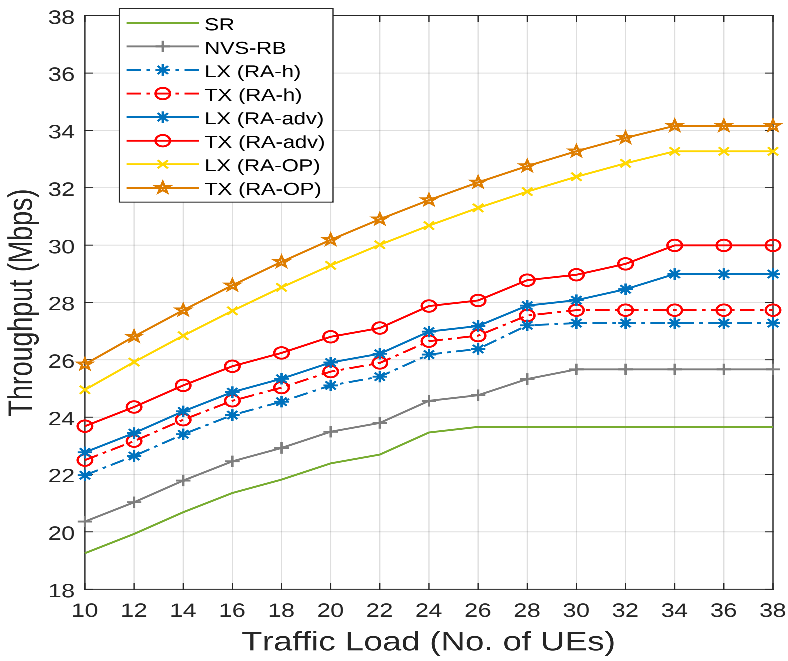
4.5.3 Optimality and Complexity Evaluation

Finally, this work presents performance comparison between optimal solutions and proposed algorithms in a small scale simulation. The simulation was conducted based on an environment including 3 SBSs, 10 to 50 UEs traffic load, 10 available PRBs in central BBU pool, totally 4 inter-tenant sharing PRBs and other unchanged parameters.

Fig.4.8(a) and Fig.4.8(b) show the system throughput of variable power allocation framework and constant power allocation framework, respectively. In both graphs, as expected optimal solutions (RPJA-OP and RA-OP) for both LX and TX show the similar developing trend as algorithms have but with higher throughput in value, which is because of the optimality and stability provided by MATLAB MILP solver. The maximum throughput improvement is still evaluated at the saturated point of all methods, in Fig.4.8(a), optimal solution LX(RPJA-OP) outperforms LX(RPJA-adv) and LX(RPJA-h) by 15.1% and 24.0% respectively; On the other hand, optimal solution TX(RPJA-OP) outperforms TX(RPJA-adv) and TX (RPJA-h) by 14.5% and 24.2% respectively. For constant power methods, Fig.4.8(b) shows that LX(RA-OP) outperforms LX(RA-adv) and LX(RA-h) by 14.5% and 19.0% respectively, as well as TX(RA-OP) outperforms LX(RA-adv) and LX(RA-h) by 11.4% and 21.8% respectively. The throughput performance compared with SR and NVS-RB is also summarized in Table 4.2.



(a) RPJA based optimization and RPJA algorithm



(b) RA based optimization and RA algorithm

Fig. 4.8 Throughput performance of MILP based optimal solution and proposed algorithms in small scale simulation

Table 4.2 AVERAGE ENHANCEMENT IN SMALL SCALE SIMULATION

Methods	Complexity	Gains on SR	Gains on NVS-RB
LX: RPJA-OP	NP-Hard	27.2%	19.5%
LX: RA-OP	NP-Hard	33.9%	25.9%
LX: RPJA-adv	$\mathcal{O}(\lvert\mathbf{I}\rvert\lvert\mathbf{C}\rvert Z)$	11.2%	4.6%
LX: RA-adv	$\mathcal{O}(\lvert\mathbf{I}\rvert\lvert\mathbf{C}\rvert Z)$	18.1%	11.0%
LX: RPJA-h	$\mathcal{O}(\lvert\mathbf{I}\rvert\lvert\mathbf{C}\rvert)$	6.8%	1.3%
LX: RA-h	$\mathcal{O}(\lvert\mathbf{I}\rvert\lvert\mathbf{C}\rvert)$	13.7%	6.9%
TX: RPJA-OP	NP-Hard	30.2%	23.3%
TX: RA-OP	NP-Hard	34.9%	29.6%
TX: RPJA-adv	$\mathcal{O}(\lvert\mathbf{I}\rvert\lvert\mathbf{C}\rvert Z)$	15.3%	8.4%
TX: RA-adv	$\mathcal{O}(\lvert\mathbf{I}\rvert\lvert\mathbf{C}\rvert Z)$	22.2%	14.9%
TX: RPJA-h	$\mathcal{O}(\lvert\mathbf{I}\rvert\lvert\mathbf{C}\rvert)$	8.9%	2.4%
TX: RA-h	$\mathcal{O}(\lvert\mathbf{I}\rvert\lvert\mathbf{C}\rvert)$	15.8%	8.9%

Table 4.3 VARIABLES AND CONSTRAINTS SIZE SCALABILITY

Problem I dimension	Variables	Constraints
M=5,N=5, $\lvert\mathbf{T}\rvert$=2, $\lvert\mathbf{P}\rvert$=3	160	1087
M=10,N=10, $\lvert\mathbf{T}\rvert$=2, $\lvert\mathbf{P}\rvert$=3	620	7272
M=10,N=50, $\lvert\mathbf{T}\rvert$=2, $\lvert\mathbf{P}\rvert$=3	3100	36192
M=50,N=50, $\lvert\mathbf{T}\rvert$=2, $\lvert\mathbf{P}\rvert$=3	15100	780352

The throughput gap between optimal solutions and proposed algorithms ranges from 15% to 25% in this small scale system but it can be reduced while scaling up the system size. Table III shows an example of the scalability. However, the complexity of proposed algorithms are simply decided by the number of iterations, users and resources (also see Table 4.3). Therefore, by sacrificing a slight loss in a large scale network scenario, the performance of both greedy and iterative algorithms is acceptable. More information about the optimality and complexity can be found in Table 4.3.

4.6 Conclusions

In this chapter, an inter-tenant resource sharing binding with variable power allocation optimization framework is proposed. The goal of the framework is to expand network capacity and user service rate by reusing the physical resources, which can be specifically realized by encouraging a flexible degree in resource sharing between tenants and allocating proper transmit power to each resource-user association. By introducing the LX and TX operations, different options are available under the optimization framework. To eliminate the curse of complexity, two resource and power joint allocation algorithms are designed and satisfying gains in terms of network performance are achieved by applying these algorithms. The improvement in networks performance has been demonstrated in the extensive numerical experiments. The total networks capacity and single user service quality can be increased by over 10% depends on different operation scenarios. In this sense, a future virtual controller with such optimization scheme can provide higher resources efficiency and network functional flexibility to all the parties in networks.

In this optimization framework, only downlink signal is considered because this work so far consider flexibly assign resources to users with aspect of networks SPs. However, with the increasing requirement in uplink, there is also need for a uplink version optimization in resource sharing and assignment as potential future extension. In this case, the uplink optimization with same resource sharing principle can be achieved by extending this framework by including more constraints, such as interference among users, various uplink service types, and potential D2D communications opportunities or restrictions. However, uplink optimization can be more complicated in networks engineering views, because users are more diverse than service provides then it would be difficult to realistically apply a resource sharing scheme to end users. The possible solution for this could be that let SPs of users to encourage users resource sharing with certain principle. In general, similar uplink optimization framework is feasible but requires more realistic sense checks than downlink optimization.

Chapter 5

SERVICE AWARE MULTI-CELL TRANSMISSION WITH INTER-TENANT SHARING IN 5G NETWORKS

Nowadays the network virtualization is brought to the spotlight due to the booming user requirements in multi-tier multi-cell wireless networks. However, as mentioned in the last chapter, the potential price for facilitating network slicing in a multi-tenant virtual network is the underutilization of treasured wireless network resources caused by varied tenant requirements and dynamic traffic episodes. A potential way to avoid such sacrifice of radio resources is to allow further inter-tenant sharing between tenants to engage a flexible reuse of resources. However, beyond this issue is the potential performance limit brought by the realistic networks themselves: the limited fronthaul capacity for data forwarding cells and different tenant service rate. Therefore, this chapter proposes a novel inter-tenant optimal resource sharing framework hand in hand with cooperative transmission (multi-cell association) technique to motivate aggressive resource utilization and solid Quality of Service (QoS) for users. In this optimization framework, the fronthaul capacity limitation is considered and resources assignment is made for satisfying certain service rate requirement from users. Meanwhile, due to the high complexity of the defined optimization problem, another iterative randomization algorithm is developed accordingly to solve the problem in polynomial time. The numerical results demonstrate significant improvement in network performance compared to certain classic resource sharing approaches.

5.1 Introduction

In emerging 5G wireless networks, clustered low power SBSs which are integrated with a MBS is deemed as the way forward in network topology for improving cell edge throughput and QoS. The SDN based network architecture in this case offers the required flexibility in network operation between SBSs and MBS by separating the control plane and user plane. However, besides these advantages, the performance of data forwarding SBSs to some extents is limited by their fronthaul links capacity, which is a physical fibre link connecting to the core network. In addition, the fronthauls usually have limited capacity for resource transmission compared to backhaul which is a pitfall in supporting high data rate. On the other hand, different service requirements from different SPs/ tenants in this case make the problem even more complicated since state of the arts at present do not manage network resources based on the dynamic varying service rate. Therefore, a novel optimization framework to both satisfy diverse service rate of tenants and assign resources taking the fronthaul load into accounts is proposed. Moreover, to achieve aggressive gain in network performance, the cooperative communications/ joint transmission technique is introduced to this framework as a powerful tool to improve poor link quality and compensate cell edge users. In terms of cooperative communications, it is one of the widely used CoMP technique. The essence of joint transmission is to allow several SBSs form a coordinated signal transmission cluster that jointly support data rate to users. Coupled with SDN based architecture, the control and measurement signal of joint transmission cells can be centrally managed by the independent control plane. In this work, the only issue constraint the performance of joint communication is the fronthaul capacity of SBSs as mentioned at the beginning.

The rest of chapter is organized as following: section 5.2 illustrates the details of optimization focuses in a network engineering horizon; In section 5.3 the problem formulations of the optimization frameworks is outlined. Besides, the realistic networks topology is explained in details; In section 5.4, an efficient iterative randomization algorithm is also proposed to provide sub-optimal solution to the problem, called IRRA algorithm and section 5.5 presents the wide sets of investigations with both large scale and small scale system simulations. At last, section 5.6 concludes the chapter.

5.2 Optimal Multi-Cell transmission and Inter-Tenant Resource Sharing

In a SDN based 5G network, the central BBU resource pool created by a independent powerful controller holds the global resources for all registered tenants and now intends to optimally partition these resources based on both tenant policy and current traffic scenario to end users. During this process, nonetheless, some issues that negatively impact the system performance come up, which includes the limited fronthaul capacity of data planes (small cells), low resource reuse efficiency caused by isolation between tenants and various service rate requirements from users. In this case, we propose an optimization framework to tackle these issues based on the following three aspects:

Fronthaul Aware Joint Cell Association

In the realistic multi-tier multi-cell networks, the UEs with different service rate requirements (tenancy dependent or independent) are randomly and non-uniformly distributed in the small cells areas at most time. In this case, the unbalanced traffic among serving small cells would result in bad signal quality or even lost of connection for certain UEs while some cells reach their fronthaul capacity. The solution to this issue could be optimally associate traffic, the UEs, by utilizing multi-point transmission and considering PRBs reuse conditions. By setting up a proper SINR threshold and applying cooperative cell association, the UE-cell association shall be made by attempting to offload traffic in extreme busy SBSs and compensate some weak quality signal links. To be noted, this aspect of proposed optimization scheme is not aiming for realizing load balancing but improving the aggregated data rate although in certain degrees the traffic load can be balanced to serve our goal. As long as the optimal data throughput and minimal lost of connection is ensured, the fronthaul aware joint cell association is deemed as a success.

Service Rate Aware Resource Slicing

Since this work is built based on a SDN enabled network, the central BBU pool is responsible for assigning the global resources (full frequency bands in network) to tenants registered in network. By applying network virtualization, such mission can be completed in a central network controller and global resources shall be sliced into different partitions for each tenants. Regarding to this, normally such resource slicing is made based on the tenancy agreement between service providers and infrastructure providers. Referred to survey in [131], there are several types of agreement: equally and statically reserve the resources for each tenant in network, dynamically partition the resources based on the proportion of cumulative traffic for each tenant and, dynamically partition the resources based on the traffic in specific cells. However, our work also aims to slice the resources based on the various service requirement/ service type of end users from different tenants rather than the amount of users in network. Therefore, our optimization framework contains constraints of rate requirement for all end users which might be tenancy dependent or independent. By assign resource in this way, the user requirement can always be satisfied dynamically.

Interference based Inter-Tenant Resource Sharing

To increase the resource reuse efficiency, we motivate an interference based inter-tenant resource sharing principle in the optimization framework to allow part of resources which is scalable in each slice/tenant to be reused by others. The inter-tenant shareable orthogonal resource blocks are dynamically decided depends on the present interference level (reused times) of them. By defining the number of inter-tenant resource blocks, efficient reuse can be achieved with a strict control of network interference level. In present state of the art, some works suggest fully reuse of all resources in BBU pool despite of the tenancy. However, from our research and by considering a congestive small cells network, there are three reasons for applying scalable inter-tenant resource sharing rather than directly reuse all resources. Firstly, fully reuse normally generates more interference in network which is hard to realized in a multi-tier small cells deployed network in realistic engineering. Secondly, the over reused resource in another way provide lower data rate for cell edge users, which results in no obvious advantage compared to scalable inter-tenant sharing in terms of network throughput and the requirement of high service rate user can be sacrificed. The last reason is that tenancy contract between infrastructure providers and tenants/operators in realistic network sometimes are strict especially when those contracts linked to operators' financial contribution which means the network operator is unable to always ask for the best resource provision in network coordination once the contracts are activated.

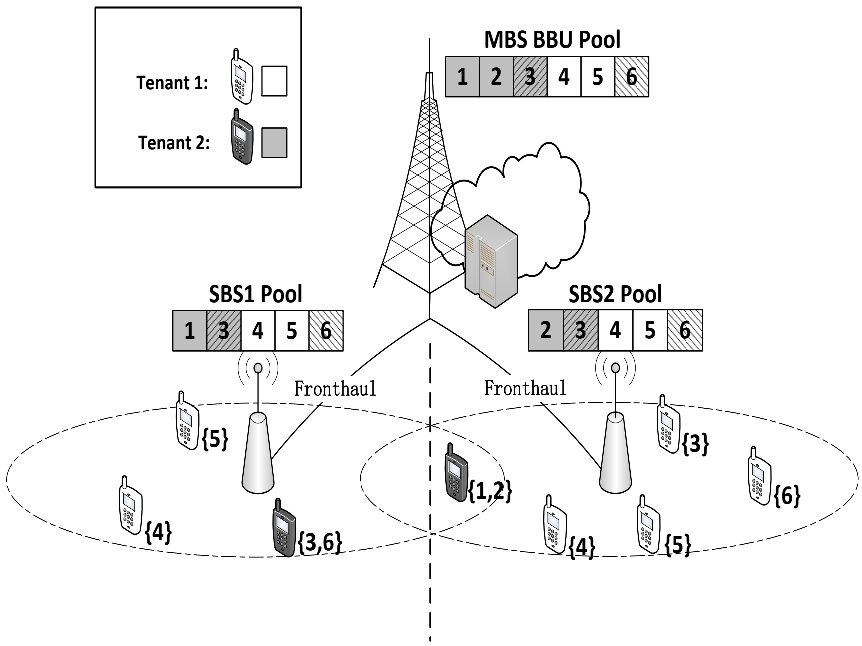


Fig. 5.1 Toy example: optimal resource reuse with multi-cell transmission. Tenant 1 users require only 1 PRBs while Tenant 2 users require 2 PRB. Limited resources in BBU pool contains 6 PRBs (3 for each tenant) and each fronthaul is limited to transmit 5 PRBs. The PRB 3 and 6 are inter-tenant sharing ones from each tenant.

In the horizon of network engineering, this optimization framework considers different aspects of realistic network scenarios. A toy example is shown in Fig.5.1 that illustrates all the principles mentioned above. In this example, the service requirement of all end users are taken into account and satisfied. Meanwhile, the only cell edge user is served by joint transmission which ensures that both cell fronthaul capacity would not be exceeded. On the other hand, the inter-tenant shareable resource blocks are defined as one of each slice resources. In such a congested environment, the inter-tenant sharing compensates the reuse limitation caused by unbalance traffic and isolate network slicing. The figure demonstrates the optimal resource allocation to users (the number marked next to user symbol) which satisfies the service requirement and generates lowest interference for reusing.

5.3 System Model and Problem Formulations

Regarding to system model, this work considers a downlink two-tier Heterogeneous network where MBSs are overlaid with (clustered) low transmit power SBSs. The MBSs are linked to C-MME, the C-GW and core packet data networks defined in the 3GPP specification via the backhauls as well as are in charge of providing control signal. On the other hand, the SBSs in so called hot zones are linked to MBSs via fronthauls. In this case, the PRBs in central BBU pool are transferred and assigned to SBSs via fronthauls. However, due to the much less capacity of fronthauls compared to backhauls, the amount of PRBs assigned to each SBS is limited and such assignment requires optimization.

The network topology is shown in Fig.5.2. The central MBS provides controlling functionalities to the whole network. Meanwhile, there are two hot zones

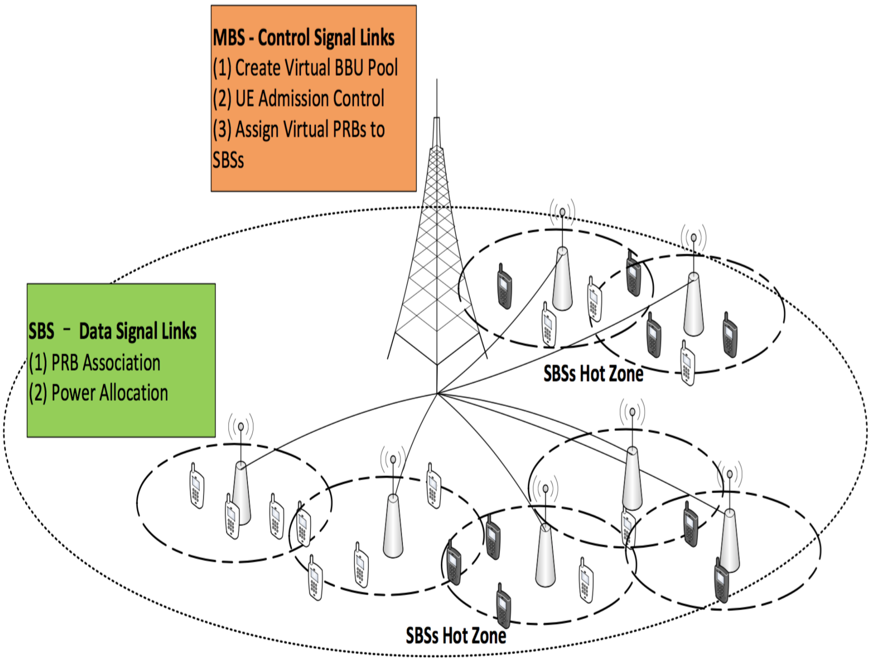


Fig. 5.2 C/U split SDN topology: intensified deployment of SBSs in a MBS control area. The traffic episode among tenants is usually non-uniformly distributed.

awaiting incoming user traffic, which will be served by low power overlapped SBSs. As mentioned, the SBSs are equipped with cooperative communication technique that up to several SBSs can serve a same cell edge user depending on certain policies setting. This topology is deemed to provide high throughput and improve cell edge user data rate. As can be seen, the MBS is responsible for creating BBU pool which centrally mange all the radio resources, controlling user admission and assign resources to SBSs based on traffic requirement. On the other hand, the SBSs are merely focusing on resource and power allocation to their connected users.

Similar to the network topology in Chapter 4, this topology is also following C/U split network architecture introduced in 3GPP release 11 [23]. The generic architecture suggests a separate control plane in various locations, however, to be in line with the topology in Chapter 4, this network topology also considers deploying the control plane into a centralized macro base station.

5.3.1 Preliminaries

In order to present the optimal cell association and resource allocation, this work defines the following binary decision variables as preliminaries.

$$x_{rt} = \begin{cases} 1 & \text{if PRB } r \text{ is assigned to tenant } t. \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

$$y_{irbt} = \begin{cases} 1 & \text{if UE } i \text{ in tenant } t \text{ associated with SBS } b \text{ uses PRB } r. \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

$$\tau_{ib} = \begin{cases} 1 & \text{if UE } i \text{ is associated to SBS } b. \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

Notations **I**, **C**, **T** and **B** are used to indicate the set of UEs, PRBs, tenants and SBSs level respectively. In this work, only mathematical set and variables are written in bold style. As mentioned before, we have $|\mathbf{C}| = M$ PRBs in resource pool and $|\mathbf{B}| = K$ SBSs (in a set **B**) waiting to serve a number of $|\mathbf{I}| = N$ users from number of $|\mathbf{T}|$ tenants, then the potential maximum power of each PRB is p_{irb}.

In this case, if users i and j are from the same tenant then there is $t = t'$.

In that respect, the SINR for a PRB-UE allocation can be estimated by the following expression,

$$\gamma_{irbt} = \frac{g_{irb} p_{irb} y_{irbt}}{\sum_{j \in I'} \sum_{d \in B'} \sum_{t' \in T} g_{ird} p_{ird} y_{jrdt'} + I_{noise}} \tag{5.4}$$

The term γ_{irbt} and I_{noise} indicates the SINR and the background noise of the channel respectively. The channel gain and transmission power are not associated with tenant therefore the t is not embedded in the definition.

Based on Eq.(5.4), the achievable data rate of UE i by taking PRB r can be approximated by the Shannon Capacity Formula,

$$R_{irbt} = \Delta f log_2(1+\gamma_{irbt}) \tag{5.5}$$

where Δf is the LTE-based frequency space for a PRB.

5.3.2 Problem Formulation

Based on the preliminaries, a joint optimization problem can be formulated as:

$$\max : \sum_{i\in\mathbf{I}}\sum_{r\in\mathbf{C}}\sum_{b\in\mathbf{B}}\sum_{t\in\mathbf{T}} R_{irbt} y_{irbt} \tag{5.6}$$

s.t.

$$\sum_{i\in\mathbf{C}}\sum_{r\in\mathbf{R}}\sum_{t\in\mathbf{T}} R_{irbt} y_{irbt} \le C_b^{fh}, \forall b \in \mathbf{B} \tag{5.7a}$$

$$\sum_{i\in\mathbf{I}}\sum_{t\in\mathbf{T}} y_{irbt} = 1, \forall r \in \mathbf{C}, \forall b \in \mathbf{B} \tag{5.7b}$$

$$\sum_{r\in\mathbf{C}} x_{rt} \ge n_t, \forall t \in \mathbf{T} \tag{5.7c}$$

$$\sum_{i\in\mathbf{I}}\sum_{b\in\mathbf{B}}\sum_{t\in\mathbf{T}} y_{irbt} \le \beta, r \in \mathbf{C} \tag{5.7d}$$

$$\sum_{r\in\mathbf{C}}\sum_{t\in\mathbf{T}} x_{rt} \le \sum_{t} n_t + \alpha \tag{5.7e}$$

$$y_{irbt} \le x_{rt}, \ \forall i \in \mathbf{I}, \forall r \in \mathbf{C}, \forall b \in \mathbf{B}, \forall t \in \mathbf{T} \tag{5.7f}$$

$$\sum_{r\in\mathbf{C}}\sum_{b\in\mathbf{B}} R_{irbt} y_{irbt} \geq R_i^{min}, \forall i \in \mathbf{I}, \forall t \in \mathbf{T} \tag{5.7g}$$

$$\sum_{r\in\mathbf{C}}\sum_{b\in\mathbf{B}} R_{irbt} y_{irbt} \leq R_i^{max}, \forall i \in \mathbf{I}, \forall t \in \mathbf{T} \tag{5.7h}$$

$$\gamma_{irbt}(y_{irbt}) \geq \gamma_{th}, \forall i \in \mathbf{I}, \forall r \in \mathbf{C}, \forall b \in \mathbf{B}, \forall t \in \mathbf{T} \tag{5.7i}$$

$$\sum_{b\in\mathbf{i}} \tau_{ib} \leq U, \forall i \in \mathbf{I} \tag{5.7j}$$

$$y_{irbt} \leq \tau_{ib}, \forall i \in \mathbf{I}, \forall r \in \mathbf{C}, \forall b \in \mathbf{B}, \forall t \in \mathbf{T} \tag{5.7k}$$

$$x_{rt} \in \{0,1\} \tag{5.7l}$$

$$y_{irtl} \in \{0,1\} \tag{5.7m}$$

$$\tau_{ib} \in \{0,1\} \tag{5.7n}$$

Constraint (5.7a) ensures that fronthaul capacity cannot be exceeded. Constraint (5.7b) indicates that each PRB in the same cell can only be associated with one UE. Furthermore, constraint (5.7c) ensures that each tenant has their basic requirement in network slicing which is confirmed by their tenancy agreement (in our case the cumulated service rate of tenant). Constraint (5.7d) indicates that every PRB can only be reused up to β times. Constraint (5.7e) restricts the sharing size α (total number of inter-tenant sharing PRBs) between tenants. Constraint (5.7f) binds two variables. Constraint (5.7g) indicates that the different service requirement of users should be satisfied while assign PRB. Constraint (5.7h) denotes that the user rate cannot exceed a certain maximum value which aims to fairly consume resources among all users. Constraint (5.7i) reflects that the SINR threshold shall be satisfied for each successful PRB-UE association. The capacity of joint transmission is indicated by constraint (5.7j) where U is the allowable number of small cells for coordinated transmission. In addition, constraint (5.7k) also binds two variables and Constraints (5.7m) and (5.7n) reassure the binary of variables.

The BBU pool links to small cells via fronthaul link which has certain capacity. In realistic network, these fronthaul link can support traffic up to several Gb/s because of the data rate encapsulation brought by common public radio interface (CPRI) protocol, which increases the dimension of data rates by controlling, managing and synchronizing the transmission signal. For instance, for a 2.5 MHZ channel transmitting signal with one antenna, the approximate user data rate it can provide is around 12 Mb/s. However, this number can be increased to 153.6 Mb/s after CPRI processing, named as CPRI data rate. The details of CPRI implementation can be found in [132]. For simplicity, this work defines the fronthaul capacity only referring to original user data rate in transmission which is much smaller than the realistic fronthaul capacity. Such assumption is also utilized by works in [133] and [134], which indicates that the ideological change in fronthaul capacity setting has no impact on the validity of optimization.

There is no doubt that the proposed optimization problem is in the family of MINLP (Mixed Integer Non-Linear Programming) problem which is difficult to solve. Therefore, we hereby design an interference threshold I_{th} instead of γ_{th} to parameterize the SINR γ_{irbt} and data rate R_{irbt} thus linearize the given problem. With the given interference threshold, the SINR estimation in Eq.(5.4) can be re-written as a parameter in Eq.(5.8) for any decided variable $y_{irbt} = 1$, to be noted, the channel gain and base station power is not tenant related.

$$\gamma_{irbt} = \frac{g_{irb} p_{irb}}{I_{th} + I_{noise}} \tag{5.8}$$

According to this, rate calculation can be also re-written as a parameter while $y_{irbt} = 1$:

$$R_{irbt} = \Delta f log_2(1 + \frac{g_{irb} p_{irb}}{I_{th} + I_{noise}}) \tag{5.9}$$

In this case, a new constraint that indicates a satisfied SINR for association is used instead of the one in Eq.(5.7i),

$$\mathbf{y_{irbt}} \sum_{j \in I'} \sum_{d \in B'} \sum_{t' \in T} g_{ird} p_{ird} y_{jrdt'} \leq I_{th}, \forall i \in \mathbf{I}, \forall r \in \mathbf{C}, \forall b \in \mathbf{B}, \forall t \in \mathbf{T} \tag{5.10}$$

As can be seen, this constraint now becomes a new non-linear constraint. Therefore, we further linearize it to the following, which is initially demonstrated by authors in [135]:

$$\begin{aligned} \sum_{j \in I'} \sum_{d \in B'} \sum_{t' \in T} g_{ird} p_{ird} y_{jrdt'} \leq I_{th} y_{irbt} + V(1 - y_{irbt}), \\ \forall i \in \mathbf{I}, \forall r \in \mathbf{C}, \forall b \in \mathbf{B}, \forall t \in \mathbf{T} \end{aligned} \tag{5.11}$$

Where V is a large enough constant number.

By combining all the transformation shown above, the linearized optimization problem now becomes the following:

$$\max : \sum_{i \in \mathbf{I}} \sum_{r \in \mathbf{C}} \sum_{b \in \mathbf{B}} \sum_{t \in \mathbf{T}} R_{irbt} y_{irbt} \tag{5.12}$$

s.t.

$$\sum_{i \in \mathbf{C}} \sum_{r \in \mathbf{R}} \sum_{t \in \mathbf{T}} R_{irbt} y_{irbt} \leq C_b^{fh}, \forall b \in \mathbf{B} \tag{5.13a}$$

$$\sum_{i \in \mathbf{I}} \sum_{t \in \mathbf{T}} y_{irbt} = 1, \forall r \in \mathbf{C}, \forall b \in \mathbf{B} \tag{5.13b}$$

$$\sum_{r\in\mathbf{C}} x_{rt} \geq n_t,\ \forall t \in \mathbf{T} \tag{5.13c}$$

$$\sum_{i\in\mathbf{I}}\sum_{b\in\mathbf{B}}\sum_{t\in\mathbf{T}} y_{irbt} \leq \beta,\ r \in \mathbf{C} \tag{5.13d}$$

$$\sum_{r\in\mathbf{C}}\sum_{t\in\mathbf{T}} x_{rt} \leq \sum_{t} n_t + \alpha \tag{5.13e}$$

$$y_{irbt} \leq x_{rt},\ \forall i \in \mathbf{I}, \forall r \in \mathbf{C}, \forall b \in \mathbf{B}, \forall t \in \mathbf{T} \tag{5.13f}$$

$$\sum_{r\in\mathbf{C}}\sum_{b\in\mathbf{B}} R_{irbt} y_{irbt} \geq R_i^{min}, \forall i \in \mathbf{I}, \forall t \in \mathbf{T} \tag{5.13g}$$

$$\sum_{r\in\mathbf{C}}\sum_{b\in\mathbf{B}} R_{irbt} y_{irbt} \leq R_i^{max}, \forall i \in \mathbf{I}, \forall t \in \mathbf{T} \tag{5.13h}$$

$$\sum_{j\in I'}\sum_{d\in B'}\sum_{t'\in T} g_{ird} p_{ird} \mathbf{y_{jrdt'}} \leq I_{th} y_{irbt} + V(1 - y_{irbt}),$$

$$\forall i \in \mathbf{I}, \forall r \in \mathbf{C}, \forall b \in \mathbf{B}, \forall t \in \mathbf{T} \tag{5.13i}$$

$$\sum_{b\in\mathbf{i}} \tau_{ib} \leq U,\ \forall i \in \mathbf{I} \tag{5.13j}$$

$$y_{irbt} \leq \tau_{ib},\ \forall i \in \mathbf{I}, \forall r \in \mathbf{C}, \forall b \in \mathbf{B}, \forall t \in \mathbf{T} \tag{5.13k}$$

$$x_{rt} \in \{0,1\} \tag{5.13l}$$

$$y_{irtl} \in \{0,1\} \tag{5.13m}$$

$$\tau_{ib} \in \{0,1\} \tag{5.13n}$$

5.3.3 Handover in Joint Transmission Enabled Architecture

The handover management and the potential overhead signalling it generates are the key issues during the process of implementing the optimization framework. Therefore, a brief illustration of the handover principle in the joint transmission enabled C/U split networks is necessary.

Introduced by [136][137][138], the widely used multi cell handover procedure normally consists of four different aspects: the serving BSs, the measurement set, the coordinating set, and the coordinating BSs. A serving BS is the one in charge of making handover decision and forwarding data to the user. A measurement set is a set of BSs whose RSRPs (Reference Signal Receiving Power) can be received and reported by the UE and feeds back to the serving BS for making the selection of next main connecting BS. A coordinating set is a set of BSs which are selected by the serving BS from the measurement set as the incoming joint communication points. Naturally, the coordinated BSs are those selected from the coordinating set by the current serving BS for sending assisting data to a UE. Without a central independent controller, the handover procedure begins by the serving BSs loading and evaluating the current information of UE mobility and power of nearby BSs. After that, the coordinating set and coordinating BSs will be selected and screened by evaluating certain RSRP threshold. At this moment, the decision of handover is made by serving BSs control plane and the selected BSs take over the UE, then the updated information of UEs and BSs is collected by new serving BSs.

However, the handover procedure is with reduced inter-cell interference and delay in the fundamental of C/U split architecture. All the handover requests and potential generated signalling are handled by central macro BS, with the BBU pool, which is fully in charge of sending measurement signal, creating set of main serving BSs and coordinating BSs and handover decision. Although there would be inter-MBS/ control plane handover procedure if several C/U split networks are considered, this issue could be optimized by proposal called inter-BBU pool handover, which can be found in research work [139]. In addition, in the numerical investigation, the simulation is conducted based on snapshot of network scenarios, which is in the form of semi-dynamic simulation. In this case, this work at present does not consider the macro BS signalling optimization, which is deemed as the future extension.

5.4 Proposed Efficient Algorithm

5.4.1 Iterative Randomized Resource Assignment Algorithm: IRRA

In this section, we propose an Iterative Randomized Resource Allocation heuristic algorithm that provides efficient sub-optimal solution in polynomial time, named IRRA.

The proposed IRRA algorithm contains two steps, shown in Algorithm 3. First step aims to iteratively (for L times) allocate and share the resource blocks to users based on the network wide interference level. The parameter L indicates a customized iterations numbers, which is integer. In order to obtain better solution from this algorithm, larger L shall be used but this also generates higher complexity in computation. More detailed can be found in next section.

In this case, we specifically define a matrix that indicates the aggregated interference in the network, $H_{N\times N}$. Besides, the mutual interfering user pairs are clearly indicated by h_{ij} elementally for a N users system.

$$H_{N\times N} = \begin{bmatrix} 0 & h_{12} & \dots & h_{1N} \\ h_{21} & 0 & \dots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} & h_{N2} & \dots & 0 \end{bmatrix} \tag{5.14}$$

At the beginning, for any given UE i, the associable PRBs for its tenancy are selected and grouped as a set $\mathbf{C_i}$, then these PRBs are sorted ascending based on their estimated determinant of the interference matrix, $\mathbf{det}(H_{N\times N})$.

Algorithm 3: Proposed IRRA Algorithm

Data: Location coordinates of UEs and SBSs; network parameters n_t, β, δ, α and γ_{th}, Interference matrix $H_{N\times N} = \emptyset$, PRBs set **C**, BSs set **B** and UE set **I**.

Result: PRBs Allocation, Cell Association and Throughput Estimation.

Step 1: Cell association and PRB allocation

for *l:=1 to L* **do**
- **for** *i:=1 to N* $\in$ **I** **do**
  - • Generate a random number ψ between 0 and 1;
  - • Obtain location and channel information of UE *i*;
  - • Mark the associable BSs set $\mathbf{B_i}$ for UE i based on distance;
  - • Mark the associable PRBs set $\mathbf{C_i}$ for UE i based on tenancy;
  - • Sort PRBs (ascending) in $\mathbf{C_i}$ based on interference matrix $H_{N\times N}$;
  - • Change sorting result of PRBs based on ψ;
  - **for** *r:=1 to M* $\in$ **C** **do**
    - **repeat**
      - • Allocate the PRB from a least used SBS;
      - **if** *all constraints satisfied* **then**
        - • y_{irbt} =1, PRB *r* is assigned to UE *i*;
        - • x_{rt} =1, Mark PRB *r* used by tenant *t*;
        - • $\tau_{ib} = 1$, UE *i* is associated to SBS *b*;
        - • Sort remained PRBs (ascending) in **C** based on updated $H_{N\times N}$;
      - **else**
        - • y_{irtl}& τ_{ib}& x_{rt} = 0, no association made;
      - **end**
    - **until** R_{min} *is exceeded*;
  - **end**
- **end**
- • $R_{(l)} = \sum R_{irbt} y_{irbt}$;

end

Step 2: Obtain the best performed throughput

• $R_{irra} = \max\{R\}$

During this process, however, a random variable $\psi \in (0,1)$ is generated to potentially change this sorting result thus the improving resource allocation and data rate output could be expected. In order to make such change with a certain possibility, a set of mathematical transformation is proposed. A parameter named critical decision value d'_r for each candidate PRB r is defined as

$$d_r = 1 - \frac{E_r}{E_{max}} \tag{5.15}$$

$$d'_r = \frac{d_r}{\sum_{r \in \mathbf{C}} d_r} \tag{5.16}$$

The E_r indicates the $\mathbf{det}(H_{N\times N})$ estimated by associating PRB r to present UE i, and E_{max} indicates the maximum $\mathbf{det}(H_{N\times N})$ among all candidate PRBs. Once d'_r for all candidate PRBs has been decided, a critical selection range for each PRB is derived as D_r. The first sorted user has critical selection range as $D_1 \in [0, d'_1]$ and $D_r \in [d'_{i-r}, d'_r + d'_{r-1}]$ is the one for the rest of PRBs. As shown in Algorithm 1, in this case the sorting result of candidate PRBs might be changed depended on the distance between ψ and each D_i. With the updated sorting result, the first candidate PRB will be associated to UE i by the least loaded SBS nearby. In this way, the fairness of choosing SBSs can be supported thus the fronthauls capacity of SBSs are difficult to exceeded in most network scenarios. It is important to note that the remain PRBs are sorted again after each association until no more association can be made due to the break clause of any optimization constraints. At the end of step 1 the sum rate for each iteration l is estimated and saved to database, and in step 2, result of the best sum rate among L Monte-Carlo iterations will be generated as the final output of proposed algorithm.

5.4.2 Complexity

However, as in the family of NP hard problem, the dimension of our optimization scheme is very high as $|\mathbf{I}|+|\mathbf{C}|+|\mathbf{B}|+|\mathbf{T}|+2|\mathbf{I}||\mathbf{T}|+|\mathbf{B}||\mathbf{C}|+3|\mathbf{I}||\mathbf{C}||\mathbf{B}||\mathbf{T}|$, which makes optimal solution impossible to achieve in realistic large size networks. With the increasing of each variables, the total dimension of the problem soars dramatically, which makes the MILP solver difficult to calculate the optimal combination for variables. On the other hand, the complexity of proposed algorithms are simply decided by the number of iterations, users and resources. As can be seen, the complexity of IRRA algorithm is $\mathscr{O}(|\mathbf{I}||\mathbf{C}|L)$. It clearly has advantage in complexity compared to the original NP-hard MILP problem. However, the solution provided by IRRA is sub-optimal, which will be discussed extensively in Numerical Investigation.

5.5 Numerical Investigation

In this section, a wide set of network level simulations is conducted based on different network scenarios. According to this, we firstly illustrate the system environment and certain assumptions that used in numerical investigation. The software defined C/U split wireless network is considered as the simulation system with a set of 3GPP standardized network parameters. Secondly, different aspects of network performance are introduced which include system throughput with multi-cell transmission optimization and with single-cell transmission optimization, cumulative density function statistic for individual UEs data rate, system throughput for different service requirement and throughput performance for different inter-tenant sharing degrees. At last, to provide a well-rounded analysis, we present a small scale system simulation that discusses the gap between MILP based optimal solution and our proposed IRRA algorithm with consideration of both complexity and scalability of the optimization problem.

Table 5.1 SIMULATION SPECIFICATION AND PARAMETERS [115]

Parameters	Values / Assumptions
Network layout	1 MBS with 6 SBSs
Cell radius (m)	MBS:1000 SBS: 200
Carrier frequency (GHz)	2
MBS antenna gain (dBi)	14
SBS antenna gain (dBi)	10
Attena configuration	1 Tx for BS, 1 Rx for UE
Thermal noise (dBm/Hz)	-174
System Bandwidth (MHz)	10
No. of PRBs in the pool	50
SBS path loss (dB)	$140.7+36.7\log_{10}(D)$ (D in km)
MBS path loss (dB)	$128.1+36.7\log_{10}(D)$ (D in km)
Shadowing standard deviation (dB)	6
Max SBS TX power (dBm)	24
MBS TX power (dBm)	43
Number of UEs	10 - 100
Max No. of a PRB reuse	6
Tenant A rate requirement (Mb/s)	0.5 - 2.5
Tenant B rate requirement (Mb/s)	2.5 - 4.5
Fronthaul capacity (Mb/s)	50

5.5.1 Simulation Scenario and Parameters

As mentioned in system model, the network topology used in numerical investigation is chosen based on software defined virtual networks. Meanwhile, the simulation environment follows the LTE-A principles proposed by 3GPP standardization in [115]. The widely used 10 MHz channel bandwidth is used in network, which indicates that a total number of 50 available PRBs can be utilized by network tenants and their UEs. Following the C/U split principle of SDN, a super MBS at the network center provides control signal to a number of standard pico cells in hot zones. The number of pico cells varies from 3 to 15 in simulation but the numerical results only shows 6 cells scenario due to the limit of pages. In addition, the location of SBSs follows the Poisson Point Process (PPP) with slightly different intensity for each simulation iteration. The assumption is made that both MBS and SBSs transmit at maximum power for signal links. Regarding to UEs, for 6 SBSs scenario, there are 10 to 100 UEs appearing in hot zones for each iteration randomly and non-uniformly. Furthermore, the UE traffic belongs to two different tenants (A and B) that tenant A is the low requirement service provider and tenant B is the high requirement provider. ratio of tenants is chosen as 1:1 but every iteration this ratio has random variation between 3:4 to 4:3. The summary of these information can be found in Table 5.1.

In order to provide well-rounded results, we compare the IRRA algorithm with two classic virtual resource slicing approaches in virtual network: resource based NVS (NVS-RB) and static reservation (SR). As mentioned in Introduction, technically, the NVS-RB approach intends to reserve a certain proportion of PRBs in BBU pool for each tenant but could automatically transfer PRBs between slices based on the traffic share between tenants. On the other hand, the SR approach holds the fixed partition of network PRBs for different tenants despite the dynamic changes in traffic. In numerical investigation, both approaches associate UEs to cells based on SINR greedy manner which always assigns UE to its nearest cells providing best signal link. The advantage of NVS-RB is the ability of dynamic adjustment for slicing resources based on incoming traffic therefore, however, both NVS-RB and SR cannot provide any inter-tenant shareable PRBs to network once the slicing process completes.

The Monte Carlo based simulation is conducted in MATLAB and the derived result data is evaluated at confidence level of 95%. The error margin of data compared to their mean value is around 2.2% (confidence interval).

5.5.2 System Evaluation

(1) Throughput Performance

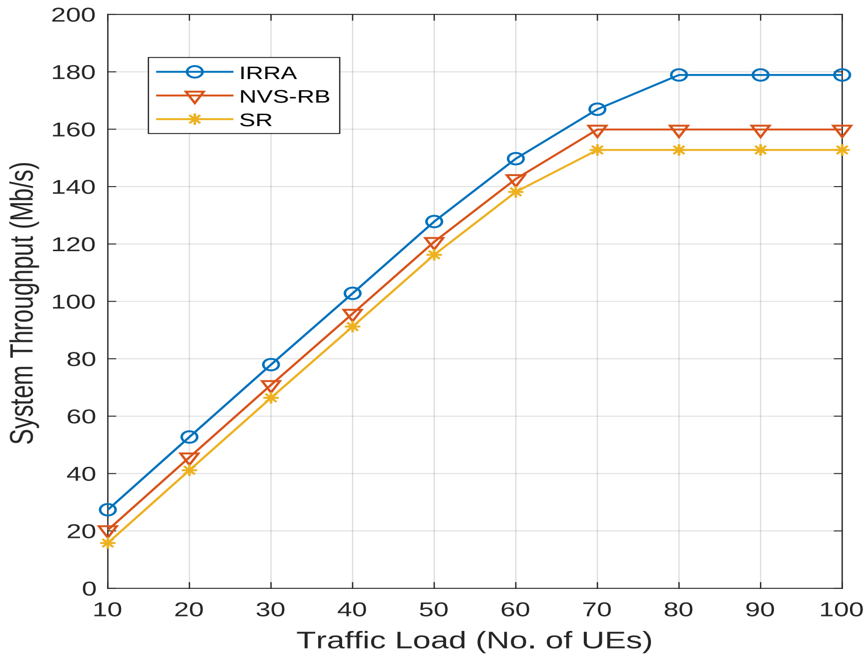


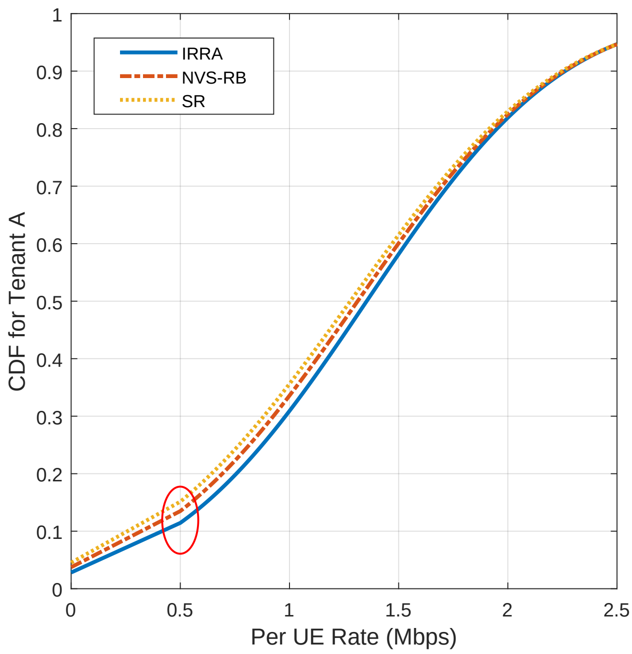
Fig. 5.3 Throughput provided by IRRA, NVS-RB and SR for different network congestion levels.

In this section, we first illustrate the system throughput performance of 6 SBSs network where maximum 3 cells joint transmission is allowed. To be noted, in the simulation we apply 10 PRBs (20% of total resources) as the inter-tenant sharing ones, the α, among two tenants in network.

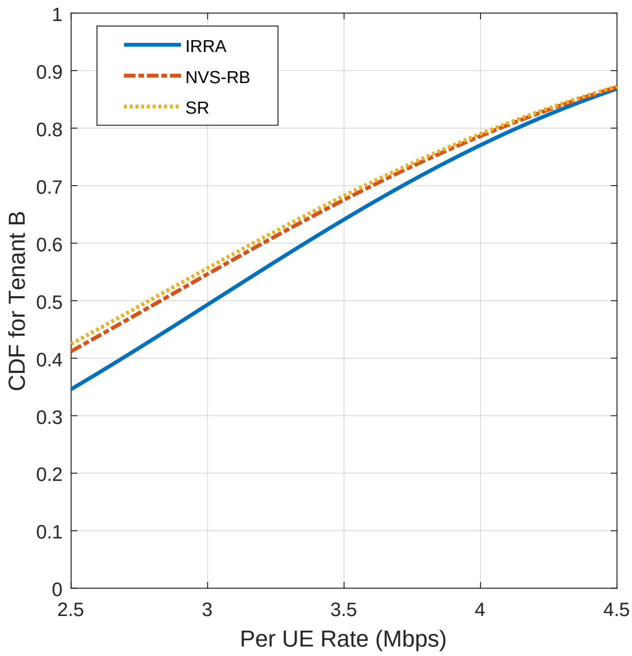
The Fig.5.3 shows the aggregated throughput of all the methods: IRRA algorithm, NVS-RB and SR. As may be observed, the total throughput offered by all the methods consistently increases with the increasing volume of UE traffic load. However, as expected the throughput at last comes to a saturated status despite which method is deployed. This happens because the aggregated UE data requirement finally exceeds the capacity of available fronthauls and the PRB reusability is at its limit. Without a doubt, the proposed IRRA algorithm is the last saturated method compared with NVS-RB and SR in the highly congested traffic scenario when network has around 80 UEs due to its fronthaul aware association and advanced inter-tenant reuse ability. The maximum throughput offered by IRRA in this case is approximately 180.9 Mb/s compared with 161.1 Mb/s and 152.8 Mb/s by NVS-RB and SR, respectively. According to this, the maximum throughput gain provided by IRRA is 12.3% and 18.4% respectively compared with NVS-RB and SR. The advantage of IRRA in this case is the more flexible PRB reuse between tenants which makes each tenant actually has higher PRB capacity for use only if the interference level is tolerant. On the other hand, the IRRA algorithm always starts associate a UE to a least used SBS which potentially considers the limit of fronthaul in those overloaded SBSs. As the result, IRRA can always provide higher throughput in a more congested network scenario than NVS-RB and SR. However, it is important to note that none of the three approaches could serve 100% traffic in in the very congested environment due to the non-uniform distribution of UEs appearance, translating to the extreme unbalanced traffic load among SBSs.

(2) Per UE Rate Performance

Apart from the aggregated system throughput performance, the performance on individual user data rate is also a meaningful indicator for evaluating resource allocation and user association method.



(a) Per UE rate performance for Tenant A



(b) Per UE rate performance for Tenant B

Fig. 5.4 Individual UE rate performance for both tenants at the most congested network load scenario of 100 UEs.

In Fig.5.4(a-b), we present the per user achievable rate based on their cumulative distribution function (CDF) statistic at the most congested load scenario: 100 UEs in the network. The X-axis of the two figures indicates the achievable data rate for the UEs and Y-axis indicates the cumulative percentage of total UEs with the data rate less than the value at X-axis. For the sake of the difference in rate requirement and presentation simplicity, the UE rate for tenant A and B is shown separately in this case. In Fig.5.4(a), as may be observed, the 0.5 Mb/s data point is where a satisfying rate for tenant A user starts to appear, which means the rate under this value is seemed as failure signal link. At the 0.5 Mb/s point, there are around 12.2% of total tenant A UEs without successful connection or satisfying rate when deploying IRRA algorithm, which is slightly lower than 14.5% and 15% offered by NVS-RB and SR, respectively. Translating that to the number of UEs, this result interprets as that there might be 2 more UEs served with their required data rate among 100 UEs thanks to the IRRA algorithm. Similarly, focusing on the 2.5 MB/s point in Fig.5.4(b), there are approximately 34.5% of Tenant B UEs cannot be served at all while applying IRRA compared to 41.8% and 42.2% by NVS-RB and SR, respectively. According to this, this result indicates that there might be up to 8 more high service rate UEs successfully served by network when IRRA is deployed. The main reason for such per UE rate improvement in both tenants is still the proper association of UEs and higher reuse capacity offered by IRRA.

(3) Throughput Variation by Different Tenant Traffic Ratio

The aim of this section is to illustrate the influence on IRRA throughput performance brought by varying tenant traffic ratio. The traffic ratio of Tenant A (low requirement service) and B (high requirement service) in this case is varied for the following: 1:2, 3:4, 4:3 and 2:1.

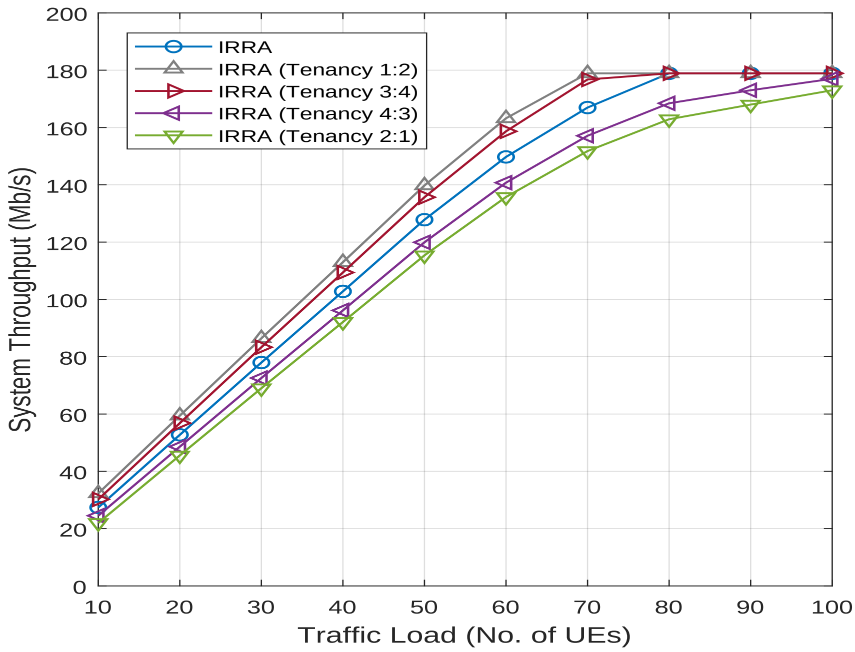


Fig. 5.5 System throughput provided by IRRA under different tenant traffic ratios and in different network congestion levels.

In Fig.5.5, the blue curve is the original 1:1 traffic ratio one for IRRA throughput. As may be observed, the IRRA deployed system throughput for all traffic ratios is increasing with the increasing load. Besides, among the different traffic ratios, the aggregated throughput for each traffic scenario before their saturated point (maximum throughput point) always ranks from highest to the lowest as, 1:2 tenant traffic ratio to 2:1 tenant traffic ratio. This is understandable since the more tenant B traffic in network means the higher aggregate throughput requirement for network. However, the capacity of PRBs and the maximum interference level in network is unchanged therefore except IRRA (2:1) all the throughput curves reach the same saturated point, which is 180.9 Mb/s.

(4) Single Cell Transmission Study Case

This section introduces the comparison between throughput performance of multi-cell transmission and single cell transmission. The goal of this study is to explore the performance limit of our proposed IRRA algorithm without the assistance of cooperative transmission.

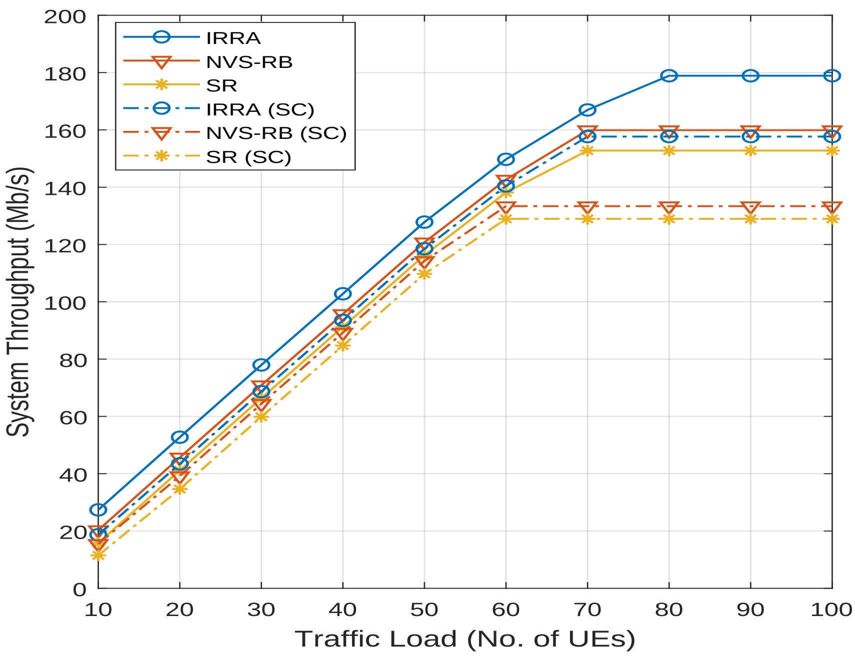


Fig. 5.6 Throughput performance comparison between multi-cell and single cell transmission.

Similarly, Fig.5.6 shows the aggregated throughput of all the methods (IRRA algorithm, NVS-RB and SR). In this case, different methods (dash curves) with 'SC' indicate the ones using single cell transmission. As expected, the throughput of all methods and transmission techniques climb with the rising of traffic load in network. However, as mey be observed, the single cell transmission throughput of all methods are all under-performed compared with those with multi-cell transmission. This indicates the value of multi-cell transmission that throughput is improved since some UEs lack of only small proportion of PRBs can usually get compensated by their cooperative cells. In our simulation, the maximum joint connection can be up to 3 SBSs but most of UEs can be served at their optimal rate by only 2 SBSs joint transmission. In addition, all three methods with single cell transmission reach their throughput saturated point earlier compare with their counterparts, which means that in the very congested network scenario multi-cell transmission can also improve the maximum throughput for network UEs. An interesting finding is that the IRRA algorithm with single cell transmission (the blue dash curve) not only outperforms NVS-RB and SR with single cell transmission but also the SR with multi-cell transmission, and nearly reach the performance of NVS-RB with multi-cell transmission. As can be seen, IRRA single cell can offer around 158.9 Mb/s system throughput which is very close to the one of multi cell NVS-RB, as 161.1 Mb/s, and is outperforming the one of single cell NVS-RB and SR, as 134.6 Mb/s and 131.1 Mb/s respectively. From the result, we can see that IRRA algorithm offers the efficient and satisfying sub-optimal resource allocation and cell association that aggressively improve system throughput with either multi-cell transmission or single cell transmission.

(5) Inter-Tenant Sharing Depth Study

In this set of numerical investigations, we study the key parameter α in our optimization scheme that potentially influence the throughput performance by adjusting the inter-tenant resource sharing depth. As mentioned at the beginning of System Evaluation, there are in total 10 PRBs can be shared between tenants in simulation. To explore the upper-bound of inter-tenant sharing, we hereby set up the range of sharing number from 0 (no sharing) to the value when no more gain appears.

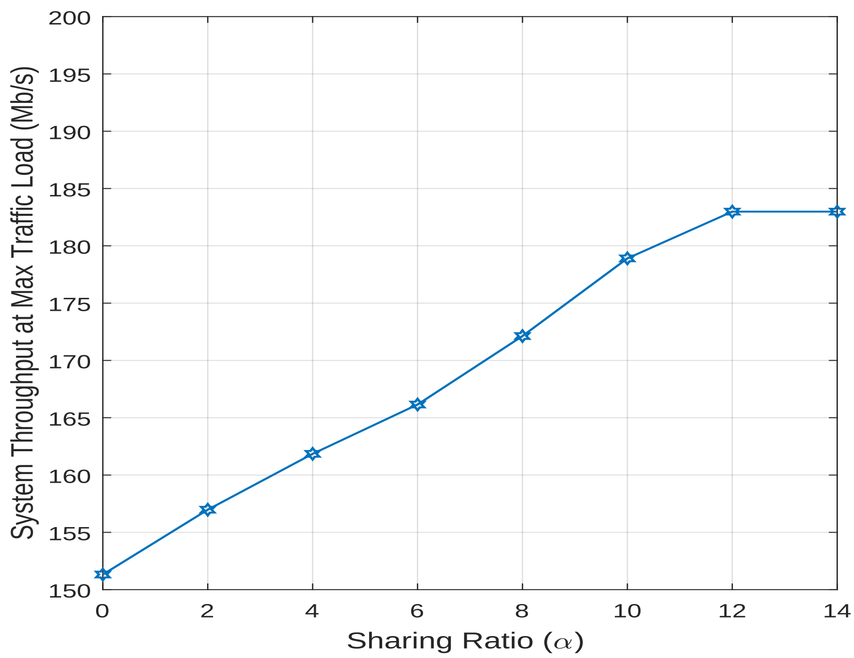


Fig. 5.7 Throughput performance comparison between multi-cell and single cell transmission.

As can be seen, in Fig.5.7, the system throughput varies with the changing of inter-tenant sharing PRBs amount at the most congested load scenario (100 UEs in network). As expected, the more available inter-tenant sharing resources in the network, the higher system throughput can be achieved. Though, finally the throughput gain stops at 12 PRBs sharing scenario as the interference level in the network is at its maximum. At this point, we can confirm that the optimal inter-tenant sharing depth for achieving maximum throughput in the 50 PRBs network environment is applying 12 PRBs. The 12 PRBs scenario achieves around 183.3 Mb/s throughput which is seemed as an aggressive 19.7% gain compared to 153.1 Mb/s achieved by no sharing scenario.

5.5.3 Optimality and Complexity Evaluation

In this section, we further present a small scale system simulation providing insight of the optimality and complexity between optimal solutions and sub-optimal solutions. The simulation was conducted based on a smaller size network topology which includes 3 SBSs, 5 to 40 UEs traffic load, 20 available PRBs, totally 6 inter-tenant sharing PRB and 15 Mb/s fronthaul capacity with other parameters unchanged in large scale system. The Fig.5.8 shows the system throughput performance of IRRA algorithm, NVS-RB, SR and MILP optimal solution. The optimal solution is calculated by MILP solver in MATLAB where a series of relaxation and branching methods are applied. The trend of throughput gain for all approaches is similar except the different saturated point of their maximum achievable throughput. In addition, we can see that IRRA starts having wider performance gap compared to optimal solution from 15 UE traffic load, which means our proposed IRRA algorithm actually is more sensitive than optimal solution in terms of rising congestion level. In summary, the total gains in maximum throughput brought by MILP optimal solution are approximately 19.9% and 23.2% compared with NVS-RB and SR, respectively. On the other hand, the same gains brought by IRRA in this simulation are approximately 11.8% and 14.9% compared

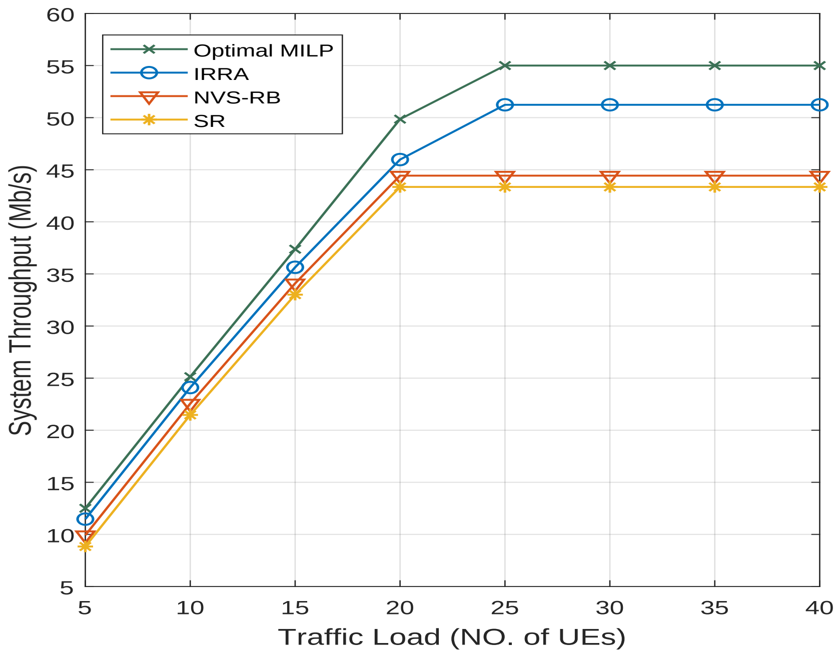


Fig. 5.8 Throughput performance comparison between multi-cell and single cell transmission.

with NVS-RB and SR. In this small scale system, the throughput performance gaps between IRRA and optimal solution are 8.1% and 8.3% for NVS-RB and SR, though this gap can be slightly varied if the system scale changes. Table 5.2 concludes the findings and Table.5.3 shows an example of the scalability.

Table 5.2 AVERAGE ENHANCEMENT IN SMALL SCALE SIMULATION

Methods	Complexity	Gains on SR	Gains on NVS-RB
Optimal	NP-Hard	23.3%	19.9%
IRRA	$\mathscr{O}(\|\mathbf{I}\|\|\mathbf{C}\|L)$	14.9%	11.8%

Besides, summarized by large scale of Monte-Carlo based simulations, the best iteration number for IRRA to find as many sub-optimal solution as it can is around 350, which takes approximately 6 - 11 Minutes for a DELL 4 Cores computer to calculate. Therefore, by sacrificing a reasonable loss in network performance,

Table 5.3 VARIABLES AND CONSTRAINTS SIZE SCALABILITY

Problem dimension	Variables	Constraints
$\|\mathbf{I}\|$=10,$\|\mathbf{C}\|$=10, $\|\mathbf{B}\|$=3, $\|\mathbf{T}\|$=2	650	1895
$\|\mathbf{I}\|$=40,$\|\mathbf{C}\|$=10, $\|\mathbf{B}\|$=3, $\|\mathbf{T}\|$=2	2540	7445
$\|\mathbf{I}\|$=40,$\|\mathbf{C}\|$=20, $\|\mathbf{B}\|$=3, $\|\mathbf{T}\|$=2	4960	14685
$\|\mathbf{I}\|$=100,$\|\mathbf{C}\|$=50, $\|\mathbf{B}\|$=6, $\|\mathbf{T}\|$=2	60700	180858

the IRRA algorithm is stable and efficient to offer sub-optimal solution to the optimization problem.

5.6 Conclusions

In this chapter, a generalized future wireless network is considered and optimized. The proposed optimization framework aims to allow more flexible inter-tenant resource sharing under a diverse service rate requirement and limited SBSs fronthauls constraints. By embedding these constraints into the optimization formulations, a well-rounded and practical wireless network can be optimized in the form of MINLP. According to this, the proposed non linear problem is further linearised by proposing a new interference threshold to the network system thus the MILP solver and an efficient iterative randomized algorithm can be applied. The numerical investigation is conducted in both large and small scale system simulation. From the large scale system result, the IRRA algorithm shows a obvious advantages in improving system throughput and per user rate compared to NVS-RB and SR approaches. On the other hand, the small scale simulation also illustrates the gap between the IRRA performance and MILP optimal solution. The trade off between optimality and operation complexity suggests that IRRA can be seemed as a solid and efficient approach for solving the optimization problem.

In terms of network engineering, this work contributes a novel solution to optimize resource assignment within a virtual network environment. Since realistic network design considerations (front-haul capacity and central BBU capacity) have been taken into account, this work can be referred to practical network engineering with further improvement. The potential improvement could be similar to the work in Chapter 4, the uplink resource sharing. As long as there is a way to easily instruct end users or their SPs to systematically re-use the spare resources, this type of development in uplink would be feasible.

Chapter 6

CONCLUSION

6.1 Summary

This thesis dives deep in 5G future wireless network architecturing by proposing a set of optimization frameworks, network planning strategies and system level embeddable algorithms. The main research gaps discovered and solved by this thesis include potential optimization in control plane load balancing, overhead signal, efficient resource sharing between network tenants and resource assignment with consideration of fronthaul engineering limitation. Literatures that provide broad view for this thesis and main contributions in Chapter 2 to 5 are reviewed as follows.

In Chapter 2, a comprehensive review on the background and related existing state of the arts and concepts with respect to the current 5G wireless network development is conducted. Popular 5G architecturing frameworks such as the cloud-RAN architecure and SDN are introduced in details. In addition, the widely used network virtualization technologies are illustrated by reviewing the development and application of virtual networks and NFV. Besides, the recent trendy state of the arts focusing on network slicing, resource scheduling and resource allocation are also reviewed as the inspiration of technical works in the thesis At last, a brief introduction of Mixed Integer Linear Programming (MILP) is presented which is applied as the fundamental modelling tool to the optimization works in this thesis.

In Chapter 3, a novel network functional decomposed topology and control plane signal optimization framework are proposed based on the present SDN architecture. The control plan load optimization focuses on optimally reducing the potential inter-SBS handover rate and balancing the control signal load in network operation. The optimization problem is formulated in the form of MILP and the performance of which indicates a satisfying reduction in handover rate compared to a classic greedy based approach. Meanwhile, the load balancing is also achieved in the numerical results compared to the same greedy approach. In addition, the trade off between handover and load balancing optimization is also investigated, which achieves a standard Pareto frontier.

In Chapter 4, by studying the current state of the arts and principles of network architecturing, an optimization framework to flexibly share and reuse the OFDMA (Orthogonal Frequency Division Multiple Access) based physical resource block (PRB) among tenants in a virtualized wireless network is proposed. In the optimization problem, two different resource sharing operations are further introduced and defined mathematically. In addition, the optimization framework is linked to advanced downlink transmit power control, in this case, extra sharing opportunities can be achiever. Due to the high complexity occurred in the problem, two efficient algorithms are designed to solve the problem without the curse of complexity. The numerical results demonstrate the system gains in throughput, individual user rate and energy efficiency.

In Chapter 5, the inter-tenant resource sharing is brought to a more practical level in terms of network engineering. The realistic wireless topology considers diverse user rate among different SPs, limited fronthaul capacity, and multi-point transmission. Based on the previous inter-tenant resource sharing framework, a novel service aware resource sharing framework with load balance is defined and linearized.

6.2 Thesis Contributions

In abstract, the main concerns in this thesis originates from difficulties in handling high demand and complexity of future 5G networks. In order to tackle these issues, a high degree of cell densification in multi-layer networks topology is deemed as mainstream in wireless networks engineering, which would result in an explosion in the utilization of various spectrum frequencies and radio resource sharing to increase the overall spatial capacity, users QoS and wireless resource usage efficiency. In that scenario, there is a need to embed quality wireless access networks optimization scheme to current networks engineering, which is the motivation for this thesis.

However, networks optimization must come with practical consideration in engineering aspect. Therefore, by reviewing recent networks engineering standards and state of the arts, research in this thesis uncovered certain research gaps and found a way to tackle them without losing generality and novelty. To link optimization to realistic networks design environment, SDN and C-RAN based architecture have been considered as the main workplace for operation. The independent control plane in these networks is the key to realize optimization because its capability in controlling a whole network from top to bottom.

Based on the current architecture design and the use of centralized controller, networks resource assignment and control functions can be realized efficiently. However, the traffic stimulated control load itself is a challenge to networks owners and users. Therefore, this thesis naturally considers a potential optimization framework that could both reduce handover signal and balance the control load among control planes.

Apart from control aspect, networks resources sharing to different service tenants is also considered as a tricky task in 5G access networks research. Regarding to this, this thesis provides an inter-tenant resource sharing optimization scheme that offers network tenants a reasonable flexibility in resource utilization or slicing. In current research, there is no such flexible inter-tenant resource sharing that can be performed by optimization computing or efficient customized algorithms. In realistic scenarios, the proposal ensures the satisfying gains in system performance by applying different scalable algorithms.

Besides, this work has been well improved in the service aware resource allocation framework in Chapter 5, where some important networks engineering constraints have been introduced. Therefore, this thesis sees the service aware resource allocation framework as the most advanced and comprehensive optimization framework from the evolution. Since this optimization scheme includes practical engineering concerns like central BBU capacity and fronthaul capacity, future 5G networks deployment would be able to embed this scheme to the networks function.

In general, these optimziation frameworks can operate in either centralized or distributed basis, depends on network topology and traffic conditions. For relative smaller networks, one single central controller would be enough to deal with all the issue centrally. However, for very large scale networks with very volatile traffic, some distributed controller can be implemented to certain hot traffic areas. In that case, distributed operation for these optimization frameworks is required. In addition to that, the variables in these optimization frameworks can come from both networks infrastructure owners (such as fronthaul capacity and BBU capacity) and service providers (such as SINR for certain services and data rate requirement). The network controller would be responsible for this type of data transmission and mutual understanding between real networks owners and virtual networks owners.

In summary, three optimization frameworks proposed in this thesis contributes a set of solutions and strategies focused on certain networks engineering related issues in 5G networks development. Based on the powerful MILP modelling, optimal solution can be achieved or expected for all frameworks, and realistic network constraints are well quantified. These optimization works can be a useful reference to 5G network engineering at the moment. In addition, the heuristic based algorithms developed in these works can also be related to certain telecommunication engineering projects, especially those working on automatic network control and planning in C-RAN or SDN.

6.3 Future Extensions From This Thesis

Although the performance of controller handover and load management scheme is satisfying in numerical investigation, there are still some issues that might require further research in future works. One issue is that there is need to test the proposed framework with more realistic data such as average overhead signal in Mb/s or Kb/s. Another issue could be the algorithms development for this proposed optimization framework. As can be seen in Chapter 3, the numerical investigation of the optimization problem is tested in a relative small scale network environment.

The potential scale of real network problem could be several times higher than the experiment size. Therefore, new efficient algorithms shall be design for this control load optimization scheme. Besides, this scheme only considers handover events between different macro controllers rather than including intra-macro handover events, hence further extension can be add more variables related to both inter and intra-macro handover to refine the framework.

The main issue occurred in resource sharing scheme is how to reduce the gap between hard to achieve optimal solutions and the algorithm based sub-optimal solutions. Further research based on this type of trade off between optimal performance and system complexity will be an interesting study. Besides, as mentioned in Chapter 4 and 5, uplink resource optimization would also be a future extensions, although uplink signals can be more complicated than downlinks. The distinguish between uplink optimization and downlink optimziation can be in terms of constraints, variables and system complexity. Research on these distinguish can be intuitive for plotting a broad understanding in 5G networks resource sharing. At last, due to these optimization frameworks operate in a semi-dynamic basis which assumes traffics stays static for a short window in experiment, more dynamic based optimization would be a massive improvement. However, this might require a complete refining in optimization models, for instance, if optimization turns into a stochastic optimization problem which adds in variables of random traffics.

References

[1] Cisco,"visual networking index", white paper, Feb. 2015. [Online]. Available: www.Cisco.com.

[2] Y. Gao, Z. Qin, Z. Feng, Q. Zhang, O. Holland, and M. Dohler, “Scalable and reliable iot enabled by dynamic spectrum management for m2m in lte-a,” *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1135–1145, Dec 2016.

[3] GSMA intelligence,"understanding 5G: perspectives on future technological advancements in mobile", white paper, 2014.

[4] S. Chen and J. Zhao, “The requirements, challenges, and technologies for 5g of terrestrial mobile telecommunication,” *IEEE Communications Magazine*, vol. 52, no. 5, pp. 36–43, May 2014.

[5] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What will 5g be?” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[6] M. Agiwal, A. Roy, and N. Saxena, “Next generation 5g wireless networks: A comprehensive survey,” *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1617–1655, thirdquarter 2016.

[7] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5g,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.

[8] J. G. Andrews, “Seven ways that hetnets are a cellular paradigm shift,” *IEEE Communications Magazine*, vol. 51, no. 3, pp. 136–144, March 2013.

[9] Y. Kishiyama, A. Benjebbour, T. Nakamura, and H. Ishii, “Future steps of LTE-A: evolution toward integration of local area and wide area systems,” *IEEE Wireless Communications*, vol. 20, no. 1, pp. 12–18, February 2013.

[10] J. Huang, R. Duan, C. Cui, and I. Chih-Lin, “Overview of cloud ran,” in *2014 XXXIth URSI General Assembly and Scientific Symposium (URSI GASS)*, Aug 2014, pp. 1–4.

[11] “C-RAN: The road towards green RAN,” *China Mobile Research Inst*, vol. 2, Feb 5 Oct. 2011.

[12] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5213–5226, Sept 2013.

[13] C. D. T. Thai, P. Popovski, M. Kaneko, and E. de Carvalho, "Multi-flow scheduling for coordinated direct and relayed users in cellular systems," *IEEE Transactions on Communications*, vol. 61, no. 2, pp. 669–678, February 2013.

[14] J. Zander and P. Mähönen, "Riding the data tsunami in the cloud: myths and challenges in future wireless access," *IEEE Communications Magazine*, vol. 51, no. 3, pp. 145–151, March 2013.

[15] S. M. A. El-atty and Z. M. Gharsseldien, "On performance of hetnet with coexisting small cell technology," in *6th Joint IFIP Wireless and Mobile Networking Conference (WMNC)*, April 2013, pp. 1–8.

[16] N. Zhang, N. Cheng, A. T. Gamage, K. Zhang, J. W. Mark, and X. Shen, "Cloud assisted hetnets toward 5g wireless networks," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 59–65, June 2015.

[17] K. M. S. Huq, S. Mumtaz, M. Alam, J. Rodriguez, and R. L. Aguiar, "Frequency allocation for hetnet comp: Energy efficiency analysis," in *ISWCS 2013; The Tenth International Symposium on Wireless Communication Systems*, Aug 2013, pp. 1–5.

[18] Z. Wang, H. Li, H. Wang, and S. Ci, "Probability weighted based spectral resources allocation algorithm in hetnet under cloud-ran architecture," in *2013 IEEE/CIC International Conference on Communications in China - Workshops (CIC/ICCC)*, Aug 2013, pp. 88–92.

[19] V. Chandrasekhar and J. G. Andrews, "Spectrum allocation in tiered cellular networks," *IEEE Transactions on Communications*, vol. 57, no. 10, pp. 3059–3068, October 2009.

[20] Y. Wu, D. Zhang, H. Jiang, and Y. Wu, "A novel spectrum arrangement scheme for femto cell deployment in lte macro cells," in *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2009, pp. 6–11.

[21] D. Lopez-Perez, G. de la Roche, A. Valcarce, A. Juttner, and J. Zhang, "Interference avoidance and dynamic frequency planning for wimax femtocells networks," in *2008 11th IEEE Singapore International Conference on Communication Systems*, Nov 2008, pp. 1579–1584.

[22] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5g multi-tier cellular wireless networks: An interference management perspective," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 118–127, June 2014.

[23] "Coordinated multipoint CoMP transmission," *3GPP LTE-Advanced Release 11*, vol. 52, no. 5, pp. 52–60, September 2011.

[24] W. Nam, D. Bai, J. Lee, and I. Kang, "Advanced interference management for 5g cellular networks," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 52–60, May 2014.

[25] R. Q. Hu and Y. Qian, "An energy efficient and spectrum efficient wireless heterogeneous network framework for 5g systems," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 94–101, May 2014.

[26] L. Sanguinetti, A. L. Moustakas, and M. Debbah, "Interference management in 5G reverse TDD HetNets with wireless backhaul: A large system analysis," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1187–1200, June 2015.

[27] K. Shen and W. Yu, "Distributed pricing-based user association for downlink heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1100–1113, June 2014.

[28] J. Xu, J. Wang, Y. Zhu, Y. Yang, X. Zheng, S. Wang, L. Liu, K. Horneman, and Y. Teng, "Cooperative distributed optimization for the hyper-dense small cell deployment," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 61–67, May 2014.

[29] Y. L. Lee, T. C. Chuah, J. Loo, and A. Vinel, "Recent advances in radio resource management for heterogeneous lte/lte-a networks," *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, pp. 2142–2180, Fourthquarter 2014.

[30] O. Galinina, S. Andreev, M. Gerasimenko, Y. Koucheryavy, N. Himayat, S. P. Yeh, and S. Talwar, "Capturing spatial randomness of heterogeneous cellular/wlan deployments with dynamic traffic," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1083–1099, June 2014.

[31] S. Talwar, D. Choudhury, K. Dimou, E. Aryafar, B. Bangerter, and K. Stewart, "Enabling technologies and architectures for 5G wireless," in *2014 IEEE MTT-S International Microwave Symposium (IMS2014)*, June 2014, pp. 1–4.

[32] "Software-defined networking the new norm for networks, ONF white paper," *Open Networking Foundation*, vol. 32, no. 6, pp. 1100–1113, April 2012.

[33] "SDN architecture overview, document ONF tr-504, ONF," 2014, pp. 7–12.

[34] A. L. L. E. Gudipati, A. Perry and Katti, "softran:," *In Proc. of ACM SIGCOMM workshop HotSDN 13. ACM, New York.*

[35] S. Sezer, S. Scott-Hayward, P. K. Chouhan, B. Fraser, D. Lake, J. Finnegan, N. Viljoen, M. Miller, and N. Rao, “Are we ready for SDN? implementation challenges for software-defined networks,” *IEEE Communications Magazine*, vol. 51, no. 7, pp. 36–43, July 2013.

[36] H. Ishii, Y. Kishiyama, and H. Takahashi, “A novel architecture for LTE-B :C-plane/U-plane split and phantom cell concept,” in *2012 IEEE Globecom Workshops*, Dec 2012, pp. 624–630.

[37] Godor, “D3.3: Final report on green network technologies,” in *INFSOICT-247733 EARTH, Tech. Rep.,*, June 2012.

[38] C. f. Lai, R. h. Hwang, H. c. Chao, M. M. Hassan, and A. Alamri, “A buffer-aware HTTP live streaming approach for SDN-enabled 5G wireless networks,” *IEEE Network*, vol. 29, no. 1, pp. 49–55, Jan 2015.

[39] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, “Design considerations for a 5G network architecture,” *IEEE Communications Magazine*, vol. 52, no. 11, pp. 65–75, Nov 2014.

[40] E. Pateromichelakis, J. Gebert, T. Mach, J. Belschner, W. Guo, N. P. Kuruvatti, V. Venkatasubramanian, and C. Kilinc, “Service-tailored user-plane design framework and architecture considerations in 5G radio access networks,” *IEEE Access*, vol. 5, pp. 17 089–17 105, 2017.

[41] P. Marsch, I. D. Silva, O. Bulakci, M. Tesanovic, S. E. E. Ayoubi, T. Rosowski, A. Kaloxylos, and M. Boldi, “5G radio access network architecture: Design guidelines and key considerations,” *IEEE Communications Magazine*, vol. 54, no. 11, pp. 24–32, November 2016.

[42] M. Y. Arslan, K. Sundaresan, and S. Rangarajan, “Software-defined networking in cellular radio access networks: potential and challenges,” *IEEE Communications Magazine*, vol. 53, no. 1, pp. 150–156, January 2015.

[43] H. H. Cho, C. F. Lai, T. K. Shih, and H. C. Chao, “Integration of SDR and SDN for 5G,” *IEEE Access*, vol. 2, pp. 1196–1204, 2014.

[44] A. Lara, A. Kolasani, and B. Ramamurthy, “Network innovation using openflow: A survey,” *IEEE Communications Surveys Tutorials*, vol. 16, no. 1, pp. 493–512, First 2014.

[45] P. W. I. F. Akyildiz and S. C. Lin, “Softair: A software defined networking architecture for 5G wireless systems,” in *Comput. Netw., vol. 85, pp. 1–18,*, 2015, pp. 7–12.

[46] B. A. A. Nunes, M. Mendonca, X. Nguyen, K. Obraczka, and T. Turletti, “A survey of software-defined networking: Past, present, and future of programmable networks,” *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1617–1634, Third 2014.

[47] "3GPP rws-120010 ws docomo, requirement, candidate solutions and technology roadmap for LTE rel-12 onward,," *3GPP Workshop on Release 12 and on-wards, Ljubljana, Slovenia*, 11-12 June 2012.

[48] L. E. Li, Z. M. Mao, and J. Rexford, "Toward software-defined cellular networks," in *2012 European Workshop on Software Defined Networking*, Oct 2012, pp. 7–12.

[49] E. J. Kitindi, S. Fu, Y. Jia, A. Kabir, and Y. Wang, "Wireless network virtualization with SDN and C-RAN for 5G networks: Requirements, opportunities, and challenges," *IEEE Access*, vol. 5, pp. 19 099–19 115, 2017.

[50] "C-RAN: The Road Towards Green RAN, 2013". [Online]. Available: http://labs.chinamobile.com/cran

[51] A. Gupta and R. K. Jha, "A survey of 5G Network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[52] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): a primer," *IEEE Network*, vol. 29, no. 1, pp. 35–41, Jan 2015.

[53] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for Mobile Networks—A Technology Overview," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

[54] Z. Cao, S. S. Panwar, M. Kodialam, and T. V. Lakshman, "Enhancing mobile networks with software defined networking and cloud computing," *IEEE/ACM Transactions on Networking*, vol. 25, no. 3, pp. 1431–1444, June 2017.

[55] C. Liu, J. Wang, L. Cheng, M. Zhu, and G. Chang, "Key microwave-photonics technologies for next-generation cloud-based radio access networks," *Journal of Lightwave Technology*, vol. 32, no. 20, pp. 3452–3460, Oct 2014.

[56] R. Wang, H. Hu, and X. Yang, "Potentials and challenges of C-RAN Supporting Multi-RATs Toward 5G Mobile Networks," *IEEE Access*, vol. 2, pp. 1187–1195, 2014.

[57] H. Zhang, Y. Dong, J. Cheng, M. J. Hossain, and V. C. M. Leung, "Fronthauling for 5G LTE-U ultra dense cloud small cell networks," *IEEE Wireless Communications*, vol. 23, no. 6, pp. 48–53, December 2016.

[58] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: insights and challenges," *IEEE Wireless Communications*, vol. 22, no. 2, pp. 152–160, April 2015.

[59] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 136–143, December 2014.

[60] R. Baldemair, T. Irnich, K. Balachandran, E. Dahlman, G. Mildh, Y. Selén, S. Parkvall, M. Meyer, and A. Osseiran, "Ultra-dense networks in millimeter-wave frequencies," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 202–208, January 2015.

[61] J. C. Juarez, A. Dwivedi, A. R. Hammons, S. D. Jones, V. Weerackody, and R. A. Nichols, "Free-space optical communications for next-generation military networks," *IEEE Communications Magazine*, vol. 44, no. 11, pp. 46–51, November 2006.

[62] T. P. McKenna, J. C. Juarez, J. A. Nanzer, and T. R. Clark, "Hybrid millimeter-wave/free-space optical system for high data rate communications," in *2013 IEEE Photonics Conference*, Sept 2013, pp. 203–204.

[63] CPRI Spec. V6.0,2013. [Online]. Available: http://www.cpri.info/

[64] F. Ponzini, L. Giorgi, A. Bianchi, and R. Sabella, "Centralized radio access networks over wavelength-division multiplexing: a plug-and-play implementation," *IEEE Communications Magazine*, vol. 51, no. 9, pp. 94–99, September 2013.

[65] J. Bartelt, P. Rost, D. Wubben, J. Lessmann, B. Melis, and G. Fettweis, "Fronthaul and backhaul requirements of flexibly centralized radio access networks," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 105–111, October 2015.

[66] D. Schulz, V. Jungnickel, C. Alexakis, M. Schlosser, J. Hilt, A. Paraskevopoulos, L. Grobe, P. Farkas, and R. Freund, "Robust optical wireless link for the backhaul and fronthaul of small radio cells," *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1523–1532, March 2016.

[67] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5G backhaul challenges and emerging research directions: A survey," *IEEE Access*, vol. 4, pp. 1743–1766, 2016.

[68] J. M. Fàbrega, M. S. Moreolo, M. Chochol, and G. Junyent, "Wdm overlay of distributed base stations in deployed passive optical networks using coherent optical ofdm transceivers," in *2012 14th International Conference on Transparent Optical Networks (ICTON)*, July 2012, pp. 1–4.

[69] Z. Ghebretensaé, K. Laraqui, S. Dahlfort, F. Ponzini, L. Giorgi, S. Stracca, J. Chen, Y. Li, J. Hansryd, and A. R. Pratt, "Transmission solutions and architectures for heterogeneous networks built as C-RANs," in *7th International Conference on Communications and Networking in China*, Aug 2012, pp. 748–752.

[70] A. Zakrzewska, S. Ruepp, and M. S. Berger, "Towards converged 5g mobile networks-challenges and current trends," in *Proceedings of the 2014 ITU kaleidoscope academic conference: Living in a converged world - Impossible without standards?*, June 2014, pp. 39–45.

[71] "wireless network cloud: Architecture and system requirements", IBM j. research and development, vol. 54, no. 1, Jan. 2010, pp. 4:1–4:12.

[72] Alcatel-lucent, "light radio white paper: Technical overview," 2013. [Online]. Available: http://www.alcatel-lucent.com/

[73] M. Gao, J. Li, D. N. K. Jayakody, H. Chen, Y. Li, and J. Shi, "A super base station architecture for future ultra-dense cellular networks: Toward low latency and high energy efficiency," *IEEE Communications Magazine*, vol. 56, no. 6, pp. 35–41, June 2018.

[74] M. Kalil, A. Al-Dweik, M. F. A. Sharkh, A. Shami, and A. Refaey, "A framework for joint wireless network virtualization and cloud radio access networks for next generation wireless networks," *IEEE Access*, vol. 5, pp. 20 814–20 827, 2017.

[75] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 358–380, Firstquarter 2015.

[76] J. S. Panchal, R. D. Yates, and M. M. Buddhikot, "Mobile network resource sharing options: Performance comparisons," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4470–4482, September 2013.

[77] M. Richart, J. Baliosian, J. Serrat, and J. Gorricho, "Resource slicing in virtual wireless networks: A survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 462–476, Sept 2016.

[78] M. Kalil, A. Shami, and Y. Ye, "Wireless resources virtualization in LTE systems," in *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2014, pp. 363–368.

[79] "RAN sharing—NEC's approach towards active radio access network sharing," in *NEC Corp.,*, 2013, pp. 363–368.

[80] B. Fan, H. Tian, and X. Yan, "A generic framework for heterogeneous wireless network virtualization: Virtual MAC design," in *2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–5.

[81] L. A. DaSilva, J. Kibiłda, P. DiFrancesco, T. K. Forde, and L. E. Doyle, "Customized services over virtual wireless networks: The path towards networks without borders," in *2013 Future Network Mobile Summit*, July 2013, pp. 1–10.

[82] B. D. M. G. . Khun-Jush, P. Bender, "Licensed shared access as complementary approach to meet spectrum demands: Benefits for next generation cellular systems," *Proc. ETSI Workshop Reconfigurable Radio Syst.*, pp. 1–7, Dec 2013.

[83] Y. Benkler, "Open wireless vs. licensed spectrum: Evidence from market adoption," *Harvard J. Law Technol.*, vol. 26, no. 1, pp. 71–163, Oct 2012.

[84] H. Zhou, R. Berry, M. L. Honig, and R. Vohra, "Complexity of allocation problems in spectrum markets with interference complementarities," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 3, pp. 489–499, March 2013.

[85] R. Berry, M. L. Honig, and R. Vohra, "Spectrum markets: motivation, challenges, and implications," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 146–155, November 2010.

[86] T. Nguyen, H. Zhou, R. A. Berry, M. L. Honig, and R. Vohra, "The impact of additional unlicensed spectrum on wireless services competition," in *2011 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, May 2011, pp. 146–155.

[87] H. Hawilo, A. Shami, M. Mirahmadi, and R. Asal, "NFV: state of the art, challenges, and implementation in next generation mobile networks (vepc)," *IEEE Network*, vol. 28, no. 6, pp. 18–26, Nov 2014.

[88] T. Taleb, "Toward carrier cloud: Potential, challenges, and solutions," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 80–91, June 2014.

[89] ETSI, "Digital cellular telecommunications system (phase 2+); universal mobile telecommunications system (umts);LTE; Network architecture," *3GPP TS 23.002 v. 11.6.0, Rel-11*, 2013.

[90] H. Lee, S. Vahid, and K. Moessner, "A survey of radio resource management for spectrum aggregation in LTE-Advanced," *IEEE Communications Surveys Tutorials*, vol. 16, no. 2, pp. 745–760, Second 2014.

[91] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June 2011.

[92] T. S. Rappaport, E. Ben-Dor, J. N. Murdock, and Y. Qiao, "38 GHz and 60 GHz angle-dependent propagation for cellular amp; peer-to-peer wireless communications," in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 4568–4573.

[93] T. O. Olwal, K. Djouani, and A. M. Kurien, "A survey of resource management toward 5G radio access networks," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1656–1686, thirdquarter 2016.

[94] H. Shajaiah, A. Abdelhadi, and C. Clancy, “Multi-application resource allocation with users discrimination in cellular networks,” in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, Sept 2014, pp. 1163–1168.

[95] M. Proebster, M. Kaschub, T. Werthmann, and S. Valentin, “Context-aware resource allocation for cellular wireless networks,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 216, Jul 2012. [Online]. Available: https://doi.org/10.1186/1687-1499-2012-216

[96] E. A. Jorswieck, “Stable matchings for resource allocation in wireless networks,” in *2011 17th International Conference on Digital Signal Processing (DSP)*, July 2011, pp. 1–8.

[97] C. Pan, C. Yin, N. C. Beaulieu, and J. Yu, “Distributed resource allocation in SDCN-Based heterogeneous networks utilizing licensed and unlicensed bands,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 711–721, Feb 2018.

[98] Y. Zhang, C. Lee, D. Niyato, and P. Wang, “Auction approaches for resource allocation in wireless systems: A survey,” *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1020–1041, Third 2013.

[99] A. Mayoral, R. Munoz, R. Vilalta, R. Casellas, R. Martinez, and V. Lopez, “Need for a transport aPI in 5G for global orchestration of cloud and networks through a virtualized infrastructure manager and planner [invited],” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A55–A62, Jan 2017.

[100] A. Rostami, A. Vidal, M. A. S. Santos, M. R. Raza, F. Moradi, B. Pechenot, Z. Ghebretensaé, P. Monti, and P. Öhlén, “An end-to-end programmable platform for dynamic service creation in 5g networks,” in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, March 2017, pp. 1–2.

[101] M. R. Raza, M. Fiorani, A. Rostami, P. Öhlen, L. Wosinska, and P. Monti, “Dynamic slicing approach for multi-tenant 5G transport networks [invited],” *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 1, pp. A77–A90, Jan 2018.

[102] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, “NVS: a substrate for virtualizing wireless resources in cellular networks,” *IEEE/ACM Transactions on Networking*, vol. 20, no. 5, pp. 1333–1346, Oct 2012.

[103] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, “Radio access network virtualization for future mobile carrier networks,” *IEEE Communications Magazine*, vol. 51, no. 7, pp. 27–35, July 2013.

[104] T. Guo and R. Arnott, “Active lte RAN sharing with partial resource reservation,” in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, Sept 2013, pp. 1–5.

[105] H. Zhang, N. Liu, X. Chu, K. Long, A. Aghvami, and V. C. M. Leung, "Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, 2017.

[106] V. Yazıcı, U. C. Kozat, and M. O. Sunay, "A new control plane for 5G network architecture with a case study on unified handoff, mobility, and routing management," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 76–85, Nov 2014.

[107] X. Yang, Y. Liu, K. S. Chou, and L. Cuthbert, "A game-theoretic approach to network slicing," in *2017 27th International Telecommunication Networks and Applications Conference (ITNAC)*, Nov 2017, pp. 1–4.

[108] S. M. A. Kazmi, N. H. Tran, T. M. Ho, and C. S. Hong, "Hierarchical matching game for service selection and resource purchasing in wireless network virtualization," *IEEE Communications Letters*, vol. 22, no. 1, pp. 121–124, Jan 2018.

[109] R. Kunst, L. Avila, E. Pignaton, S. Bampi, and J. Rochol, "A resources sharing architecture for heterogeneous wireless cellular networks," in *2016 IEEE 41st Conference on Local Computer Networks (LCN)*, Nov 2016, pp. 228–231.

[110] I. Malanchini, S. Valentin, and O. Aydin, "Generalized resource sharing for multiple operators in cellular wireless networks," in *2014 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Aug 2014, pp. 803–808.

[111] H. P. Williams and S. C. Brailsford, "Advances in linear and integer programming," J. E. Beasley, Ed. New York, NY, USA: Oxford University Press, Inc., 1996, ch. Computational Logic and Integer Programming, pp. 249–281. [Online]. Available: http://dl.acm.org/citation.cfm?id=247975.247983

[112] A. Bemporad and M. Morari, "Control of systems integrating logic, dynamics, and constraints," *Automatica*, vol. 35, no. 3, pp. 407 – 427, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005109898001782

[113] C. A. Floudas, "Nonlinear and mixed-integer programming – fundamentals and applications," *Oxford University Press*, 1995.

[114] Matlab, "Mixed-integer linear programming solver," *MathWorks*. [Online]. Available: https://uk.mathworks.com/help/optim/ug/intlinprog.html

[115] G. R.-. W. Docomo, "Requirement, candidate solutions and technology roadmap for lte rel-12 onward,d, 3GPP workshop on release 12 and onwards ljubljana, slovenia," 11-12 June 2012.

[116] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 82–89, February 2014.

[117] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli, "Control-data separation architecture for cellular radio access networks: A survey and outlook," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 446–465, Firstquarter 2016.

[118] Z. R. Zaidi, V. Friderikos, O. Onireti, J. Gang, and M. A. Imran, *An Integrated Approach for Functional Decomposition of Future RAN*. Cham: Springer International Publishing, 2016, pp. 123–144. [Online]. Available: https://doi.org/10.1007/978-3-319-27568-0_6

[119] "3rd generation partnership project:3GPP," no. TR 36.814 V9.0.0, 2010.

[120] F. R. Colazzo, A. and Lambiase, "Achieving low-latency communication in future wireless networks: the 5G NORMA approach," *5GPPP Architecture Working Group, View on 5G Architecture, White paper*, Dec 2017.

[121] E. N. ISG, "Gs nfv-eve 005 v1.1.1 network function virtualisation (nfv); ecosystem; report on sdn usage in nfv architectural framework," *ETSI V1.1.1*, Dec, 2015.

[122] T. K. Thuc, E. Hossain, and H. Tabassum, "Downlink power control in two-tier cellular networks with energy-harvesting small cells as stochastic games," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 5267–5282, Dec 2015.

[123] D. Qin, W. Xu, and Z. Ding, "Power control and resource allocation for capacity improvement in picocell downlinks," in *2012 International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct 2012, pp. 1–6.

[124] C. Vlachos and V. Friderikos, "MOCA: multiobjective cell association for device-to-device communications," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9313–9327, Oct 2017.

[125] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 30–37, June 2011.

[126] C. Bae and W. E. Stark, "Energy-bandwidth tradeoff with spatial reuse in wireless multi-hop networks," in *MILCOM 2008 - 2008 IEEE Military Communications Conference*, Nov 2008, pp. 1–7.

[127] A. Chockalingam and M. Zorzi, "Energy efficiency of media access protocols for mobile data networks," *IEEE Transactions on Communications*, vol. 46, no. 11, pp. 1418–1421, Nov 1998.

[128] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 11, pp. 3874–3886, November 2011.

[129] H. Mahdavi-Doost, N. Prasad, and S. Rangarajan, "Optimizing energy efficiency over energy-harvesting LTE cellular networks," *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 3, pp. 320–332, Sept 2017.

[130] R. Mahapatra, Y. Nijsure, G. Kaddoum, N. U. Hassan, and C. Yuen, "Energy efficiency tradeoff mechanism towards wireless green communication: A survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 686–705, Firstquarter 2016.

[131] P. Caballero, A. Banchs, G. de Veciana, and X. Costa-Pérez, "Multi-tenant radio access network slicing: Statistical multiplexing of spatial loads," *IEEE/ACM Transactions on Networking*, vol. 25, no. 5, pp. 3044–3058, Oct 2017.

[132] A. de la Oliva, J. A. Hernández, D. Larrabeiti, and A. Azcorra, "An overview of the CPRI specification and its application to C-RAN-based LTE scenarios," *IEEE Communications Magazine*, vol. 54, pp. 152–159, 2016.

[133] Y. L. Lee, J. Loo, T. C. Chuah, and A. A. El-Saleh, "Fair resource allocation with interference mitigation and resource reuse for LTE/LTE-A femtocell networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 8203–8217, Oct 2016.

[134] Y. L. Lee, J. Loo, T. C. Chuah, and L. Wang, "Dynamic network slicing for multitenant heterogeneous cloud radio access networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2146–2161, April 2018.

[135] B. Niu, Y. Zhou, H. Shah-Mansouri, and V. W. S. Wong, "A dynamic resource sharing mechanism for cloud radio access networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8325–8338, Dec 2016.

[136] X. Z. Cheng-Chung Lin, Kumbesan Sandrasegaran and Z. Xu, "Limited CoMP handover algorithm for lte-advanced," *Journal of Engineering*, vol. 2013, Feb 2013.

[137] R. Shrivastava and M. C. Aguayo-Torres, "Analysis and evaluation of cooperative multi-point transmission/reception and soft handover for LTE-Advanced," in *2012 World Congress on Information and Communication Technologies*, Oct 2012, pp. 826–830.

[138] R. Y. Kim, I. Jung, X. Yang, and C. Chou, "Advanced handover schemes in IMT-advanced systems [WiMAX/LTE Update]," *IEEE Communications Magazine*, vol. 48, no. 8, pp. 78–85, August 2010.

[139] H. Zhang, C. Jiang, J. Cheng, and V. C. M. Leung, "Cooperative interference mitigation and handover management for heterogeneous cloud small cell networks," *IEEE Wireless Communications*, vol. 22, no. 3, pp. 92–99, June 2015.